By the same author

Intention

An Introduction to Wittgenstein's Tractatus

Three Philosophers (with Peter Geach)

THE COLLECTED PHILOSOPHICAL PAPERS OF G. E. M. ANSCOMBE

VOLUME TWO

Metaphysics and the Philosophy of Mind

Basil Blackwell · Oxford

First published in 1981 by
Basil Blackwell Publisher
108 Cowley Road
Oxford OX4 1JF
England

British Library Cataloguing in Publication Data

Anscombe, Gertrude Elizabeth Margaret
The collected philosophical papers of G. E. M. Anscombe
Vol. 2: Metaphysics and the philosophy of mind
1. Philosophy, English – Addresses, essays, lectures
I. Title
192'.08 B1618

ISBN 0-631-12932-4

Typeset in Photon Baskerville

Contents

Introduction vii

PART ONE: The Philosophy of Mind

1 The Intentionality of Sensation: A Grammatical Feature 3
2 The First Person 21
3 Substance 37
4 The Subjectivity of Sensation 44
5 Events in the Mind 57
6 Comments on Professor R. L. Gregory's Paper on Perception 64
7 On Sensations of Position 71
8 Intention 75
9 Pretending 83
10 On the Grammar of 'Enjoy' 94

PART TWO: Memory and the Past

11 The Reality of the Past 103
12 Memory, 'Experience' and Causation 120

PART THREE: Causality and Time

13 Causality and Determination 133
14 Times, Beginnings and Causes 148
15 Soft Determinism 163
16 Causality and Extensionality 173
17 Before and After 180
18 Subjunctive Conditionals 196
19 "Under a Description" 208
20 *Analysis* Competition – Tenth Problem 220
21 A Reply to Mr C. S. Lewis's Argument that "Naturalism" is Self-Refuting 224

Index 233

Introduction

My first strenuous interest in philosophy was in the topic of causality. I didn't know that what I was interested in belonged to philosophy. As a result of my teen-age conversion to the Catholic Church – itself the fruit of reading done from twelve to fifteen – I read a work called *Natural Theology* by a nineteenth-century Jesuit. I read it with great appetite and found it all convincing except for two things. One was the doctrine of *scientia media*, according to which God knew what anybody would have done if, e.g., he hadn't died when he did. This was a part of theodicy, and was also the form in which the problem of counter-factual conditionals was discussed. I found I could not believe this doctrine: it appeared to me that there was not, quite generally, any such thing as what would have happened if what did happen had not happened, and that in particular there was no such thing, generally speaking, as what someone would have done if . . . and certainly that there was no such thing as how someone would have spent his life if he had not died a child. I did not know at the time that the matter was one of fierce dispute between the Jesuits and the Dominicans, who took rather my own line about it. So when I was being instructed a couple of years later by a Dominican at Oxford, Fr Richard Kehoe, and he asked me if I had any difficulties, I told him that I couldn't see how that stuff could be true. He was obviously amused and told me that I certainly didn't have to believe it, though I only learned the historical fact I have mentioned rather later.

But it was the other stumbling block that got me into philosophy. The book contained an argument for the existence of a First Cause, and as a preliminary to this it offered a proof of some 'principle of causality' according to which anything that comes about must have a cause. The proof had the fault of proceeding from a barely concealed assumption of its own conclusion. I thought that this was some sort of carelessness on the part of the author, and that it just needed tidying up. So I started writing improved versions of it; each one satisfied me for a time, but then reflection would show me that I had committed the same fault. I don't think I ever showed my efforts to anyone; I tore them up when I found they were no good, and I went round asking people *why*, if something happened, they would be sure it had a cause. No one had an answer to this. In two or three years of effort I produced five versions of a would-be proof, each one of which I then found guilty of the same error, though each time it was more cunningly concealed. In all this time I had no philosophical teaching about the matter; even my last attempt was made before I started reading Greats at Oxford. It was not until then that I read Hume and the discussion in Aquinas, where he says that it isn't part of the concept of *being* to include any relation to a cause. But I could not understand the grounds of his further claim, that it *is* part of the concept of *coming into being*.

The other central philosophical topic which I got hooked on without even realizing that it was philosophy, was perception. I read a book by Fr Martin D'Arcy, S. J., called *The Nature of Belief* and got just that out of it. I was sure that I saw objects, like packets of cigarettes or cups or . . . any more or less substantial thing would do. But I think I was concentrated on artefacts, like other products of our urban life, and the first more natural examples that struck me were 'wood' and the sky. The latter hit me amidships because I was saying dogmatically that one must know the category of object one was speaking of – whether it was a colour or a kind of stuff, for example; *that* belonged to the logic of the term one was using. It couldn't be a matter of empirical discovery that something belonged to a different category. The sky stopped me.

For years I would spend time, in cafés, for example, staring at objects saying to myself: 'I see a packet. But what do I really see? How can I say that I see here anything more than a yellow expanse?' While still doing Honour Mods, and so not yet having got into my undergraduate philosophy course, I went to H. H. Price's lectures on perception and phenomenalism. I found them intensely interesting. Indeed, of all the people I heard at Oxford, he was the one who excited my respect; the one I found worth listening to. This was not because I agreed with him, indeed, I used to sit tearing my gown into little strips because I wanted to argue against so much that he said. But even so, what he said seemed to me to be absolutely about the stuff. The only book of his that I found so good was *Hume's Theory of the External World* which I read straight on from first sentence to last. Again, I didn't agree with some of it; he offered an amended account of identity to rewrite Hume, in a way that seemed to me to miss the force of Hume's thoughts about identity as seeming to be "midway betwixt unity and diversity": he wanted to amend Hume into starting with the idea that identity really belonged just to atomic sense-impressions – which won't work because "every sense-impression contains temporal parts"; and then changing to the conception of "identical" as applying always to a *whole*, having temporal parts or spatial parts or both, and never to a single indivisible entity, if such there be. That is, he wanted to smooth Hume out. But he was really writing about the stuff itself, even if one did not accept his amendment. It was he who had aroused my intense interest in Hume's chapter "On scepticism with regard to the senses".

I always hated phenomenalism and felt trapped by it. I couldn't see my way out of it but I didn't believe it. It was no good pointing to difficulties about it, things which Russell found wrong with it, for example. The strength, the central nerve of it remained alive and raged achingly. It was only in Wittgenstein's classes in 1944 that I saw the nerve being extracted, the central thought "I have got *this*, and I define 'yellow' (say) as *this*" being effectively attacked. – At one point in these classes Wittgenstein was discussing the interpretation of the sign-post, and it burst upon me that the way you go by it is the final interpretation. At another I came out with "But I still want to say: Blue is there." Older hands smiled or laughed but Wittgenstein checked

them by taking it seriously, saying "Let me think what medicine you need. . . . Suppose that we had the word 'painy' as a word for the property of some surfaces." The 'medicine' was effective, and the story illustrates Wittgenstein's ability to understand the thought that was offered to him in objection. One might protest, indeed, that there is *this* wrong with Locke's assimilation of secondary qualities to pain: you can sketch the functioning of "pain" as a word for a secondary quality, but you can't do the reverse operation. But the 'medicine' did not imply that you could. If "painy" were a possible secondary quality word, then wouldn't just the same motive drive me to say: "Painy is there" as drove me to say "Blue is there"? I did not mean "'Blue' is the name of *this* sensation which I am having," nor did I switch to that thought.

This volume contains the earliest purely philosophical writing on my part which was published: the criticism of C. S. Lewis' argument for 'the self-refutation of the Naturalist' in the first edition of his book, *Miracles*, chapter III. Those who want to see what the argument was, without relying on my criticism for it, should take care to get hold of the first edition (1947). The version of that chapter which is most easily available is the second edition, which came out as a Fontana paperback in 1960. The chapter, which in 1947 had the title "The Self-Contradiction of the Naturalist", was rewritten and is now called "The Cardinal Difficulty of the Naturalist". The last five pages of the old chapter have been replaced by ten pages of the new, though a quotation from J. B. S. Haldane is common to both. Internal evidence shows that at least some of the rewriting was done after the first Sputnik and even after the hot dry summer of 1959. But I should judge that he thought rather hard about the matter in the interval. The rewritten version is much less slick and avoids some of the mistakes of the earlier one; it is much more of a serious investigation. He distinguishes between 'the Cause–Effect *because*' and 'the Ground–Consequent *because*', where before he had simply spoken of 'irrational causes'. If what we think at the end of our reasoning is to be true, the correct answer to "Why do you think that?" must use the latter *because*. On the other hand, every event in Nature must be connected with previous events in the Cause-and-Effect relation. . . . "Unfortunately the two systems are wholly distinct". . . . And "even if grounds do exist, what exactly have they got to do with the actual occurrence of the belief as a psychological event?"

These thoughts lead him to suggest that being a cause and being a proof must coincide – but he finds strong objections to this. (He obviously had imbibed some sort of universal-law determinism about causes.) After some consideration he reverts to the (unexamined) idea he used in the first edition, of 'full explanation': "Anything which professes to explain our reasoning fully without introducing an act of knowing, thus solely determined by what is known, is really a theory that there is no reasoning. But this, as it seems to me, is what Naturalism is bound to do." The remaining four and a half pages are devoted to an elaboration of this. Unluckily he doesn't explore this idea

of 'an act of knowing solely determined by what is known', which is obviously crucial.

Rereading the argument of the first edition and my criticisms of it, it seems to me that they are just. At the same time, I find them lacking in any recognition of the depth of the problem. I don't think Lewis' first version itself gave one much impression of that. The argument of the second edition has much to criticize in it, but it certainly does correspond more to the actual depth and difficulty of the questions being discussed. I think we haven't yet an answer to the question I have quoted from him: "What is the connection between grounds and the actual occurrence of the belief?"

The fact that Lewis rewrote that chapter, and rewrote it so that it now has these qualities, shows his honesty and seriousness. The meeting of the Socratic Club at which I read my paper has been described by several of his friends as a horrible and shocking experience which upset him very much. Neither Dr Havard (who had Lewis and me to dinner a few weeks later) nor Professor Jack Bennett remembered any such feelings on Lewis' part. The paper that I read is as printed here. My own recollection is that it was an occasion of sober discussion of certain quite definite criticisms, which Lewis' rethinking and rewriting showed he thought were accurate. I am inclined to construe the odd accounts of the matter by some of his friends – who seem not to have been interested in the actual arguments or the subject-matter – as an interesting example of the phenomenon called "projection".

Part One

The Philosophy of Mind

1 The Intentionality of Sensation

A Grammatical Feature

I *Intentional Objects*

Berkeley calls "colours with their variations and different proportions of light and shade" the "proper" and also the "immediate" objects of sight.[1] The first at any rate long seemed obvious to everyone, both before Berkeley and since his time. But Berkeley's whole view is now in some disrepute. Sense-data, a thoroughly Berkeleyan conception given that name by Russell, have become objects of ridicule and contempt among many present-day philosophers.

That word "object" which comes in the phrase "object of sight" has suffered a certain reversal of meaning in the history of philosophy, and so has the connected word "subject", though the two reversals aren't historically connected. The subject used to be what the proposition, say, is about: the thing itself as it is in reality – unprocessed by being conceived, as we might say (in case there is some sort of processing there); objects on the other hand were formerly always objects *of* —. Objects of desire, objects of thought, are not objects in one common modern sense, not indivdual things, such as *the objects found in the accused man's pockets*.

I might illustrate the double reversal by a true sentence constructed to accord with the *old* meanings: subjectively there must be some definite number of leaves on a spray that I see, but objectively there need not: that is, there need not be some number such that I *see* that number of leaves on the spray.

When Descartes said that the cause of an idea must have at least as much formal reality as the idea had objective reality, he meant that the cause must have at least as much to it as what the idea was of would have, if what the idea was of actually existed. The "*realitas objectiva*" of an idea thus meant what we should call its "content" – namely what it is of, but considered as belonging purely to the idea. "What a picture is of" can easily be seen to have two meanings: what served as a model, what the picture was taken from – and what is to be seen in the picture itself, which may not even have had an original.

Thus formerly if something was called an object that would have raised the question "object of what?" It is hardly possible to use the word "object" in

[1] Throughout this paper I use double quotes for ordinary quotations (and so singles for quotes within quotes) and singles I use as scare quotes.

From R. J. Butler (ed.), *Analytical Philosophy*, second series (Oxford, 1965).

this way nowadays unless it actually occurs in such a phrase as "object of desire" or "object of thought". Suppose somebody says that the object of desire, or desired object, need not exist, and so there need not be any object which one desires. He is obviously switching from one use of the word "object" to another. If, however, we speak of objects of sight, or seen objects, it will usually be assumed that "objects" has the more modern sense: these will be objects, things, entities, which one sees. Now to prevent confusion I will introduce the phrase "intentional object" to mean "object" in the older sense which still occurs in "object of desire".

"Intentional" in these contexts is often spelt with an s. This was an idea of Sir William Hamilton's; he wanted to turn the old logical word "intention" into one that looked more like "extension". I prefer to keep the older spelling with two ts. For the word is the same as the one in common use in connection with action. The concept of intention which we use there of course occurs also in connection with *saying*. That makes the bridge to the logician's use.

There are three salient things about intention which are relevant for my subject. First, not any true description of what you do describes it as the action you intended: only under certain of its descriptions will it be intentional. ("Do you mean to be using that pen?" – "Why, what about this pen?" – "It's Smith's pen." – "Oh Lord, no!") Second, the descriptions under which you intend what you do can be vague, indeterminate. (You mean to put the book down on the table all right, and you do so, but you do not mean to put it down anywhere in particular on the table – though you do put it down somewhere in particular.) Third, descriptions under which you intend to do what you do may not come true, as when you make a slip of the tongue or pen. You act, but your intended act does not happen.

Intentionality, whose name is taken from intention and expresses these characteristics of the concept *intention*, is found also in connection with many other concepts. I shall argue that among these are concepts of sensation. Like many concepts marked by intentionality, though unlike intention itself, these are expressed by verbs commonly taking direct objects. I shall speak of intentional verbs, taking intentional objects. I have mentioned the history of the word "object" to forestall any impression that "an intentional object" means "an intentional entity".

Obvious examples of intentional verbs are "to think of", "to worship", "to shoot at". (The verb "to intend" comes by metaphor from the last – "*intendere arcum in*", leading to "*intendere animum in*".) Where we have such a verb taking an object, features analogous to the three features of intentionalness in action relate to some descriptions occurring as object-phrases after the verb.

The possible non-existence of the object, which is the analogue of the possible non-occurrence of the *intended* action, is what has excited most attention about this sort of verb. "Thinking of" is a verb for which the topic of the non-existent object is full of traps and temptations; "worshipping" is

less dangerous and may help us to keep our heads. Consider the expression "object of thought". If I am thinking of Winston Churchill then he is the object of my thought. This is like "What is the object of these people's worship?" Answer: "The moon." But now suppose the object of my thought is Mr Pickwick, or a unicorn; and the object of my worship is Zeus, or unicorns. With the proper names I named no man and no god, since they name a fictitious man and a false god. Moreover Mr Pickwick and Zeus are nothing but a fictitious man and a false god (contrast the moon, which, though a false god, is a perfectly good heavenly body). All the same it is clear that "The Greeks worshipped Zeus" is true. Thus "*X* worshipped —" and "*X* thought of —" are not to be assimilated to "*X* bit —". For, supposing "*X*" to be the name of a real person, the name of something real has to be put in the blank space in "*X* bit —" if the completed sentence is to have so much as a chance of being true. Whereas in "*X* worshipped —" and "*X* thought of —" that is not so.

This fact is readily obscured for us because with "*X* thought of —" the more frequent filling-in of the blank is a name or description of something real; for when the blank is filled in so in a true sentence, it is the real thing itself, not some intermediary, that *X* thought of. This makes it look as if the reality of the object mattered, as it does for biting. Nevertheless, it is obvious that vacuous names can complete such sentence-frames. So perhaps they stand in such frames for something with a *sort* of reality. That is the hazy state of mind one may be in about the matter.

A not very happy move to clarify it is to say, "Well, *X* had his idea of Zeus, or unicorns, or Mr Pickwick, and that gives you the object you want." This is an unhappy move on several counts. First, it makes it seem that the *idea* is what *X* was worshipping or thinking of. Second, the mere fact of real existence (is this now beginning to be opposed to existence of some other kind?) can't make so very much difference to the analysis of a sentence like "*X* thought of —". So if the idea is to be brought in when the object doesn't exist, then equally it should be brought in when the object does exist. Yet one is thinking, surely, of Winston Churchill, not of the idea of him, and just that fact started us off. When one reads Locke, one wants to protest: "The mind is not employed about ideas, but about things – unless ideas are what we happen to be thinking about." Whatever purpose is served by introducing ideas, by saying, "Well, they had an idea of Zeus," we cannot say that the idea is the object of thought, or worship. It will not be right to say *X* worshipped an idea. It is rather that the subject's having an idea is what is needed to give the proposition a chance of being true. This may seem helpful for "worshipping", but not for "thinking of"; "thinking of" and "having an idea of" are too similar; if the one is problematic, then so is the other.

Let us concentrate on the fact that many propositions containing intentional verbs are true, and let us not be hypnotized by the possible non-existence of the object. There are other features too: non-substitutability of different descriptions of the object, where it does exist; and possible indeter-

minacy of the object. In fact all three features are connected. I can think of a man without thinking of a man of any particular height; I cannot hit a man without hitting a man of some particular height, because there is no such thing as a man of no particular height. And the possibility of this indeterminacy makes it possible that when I am thinking of a particular man, not every true description of him is one under which I am thinking of him.

I will now define an intentional verb as a verb taking an intentional object; intentional objects are the sub-class of direct objects characterized by these three connected features. By this definition, "to believe" and "to intend" are not themselves intentional verbs, which may seem paradoxical. But, say, "to believe – to be a scoundrel" will accord with the definition, so that it is not so paradoxical as to leave out belief and intention altogether.

But now comes a question: ought we really to say that the intentional object is a bit of language, or may we speak as if it were what the bit of language stands for? As grammarians and linguists use the words nowadays "direct object" and "indirect object" stand for parts of sentences. So if I call intentional objects a sub-class of direct objects, that may seem already to determine that an intentional object is a bit of language.

However, the matter is not so easily settled. Of course I do not want to oppose the practice of grammarians. But it is clear that the concept of a direct object – and hence the identification of the sentence-part now called the direct object – is learned somewhat as follows: the teacher takes a sentence, say "John sent Mary a book" and says: "What did John send Mary?" Getting the answer "A book" he says: "That's the direct object." Now the question does not really suppose, and the pupil, if he goes along with the teacher, does not take it, that any particular people, of whom the sentence is true, are in question, and so we may say that when the teaching is successful the question is understood as equivalent to "What does the sentence 'John sent Mary a book' say John sent Mary?" The grammatical concept of a direct object is acquired by one who can answer any such question. The correct answer to such a question gives (in older usage) or itself is (in more recent usage) the direct object. Now suppose that someone were to ask: "What is communicated to us by the phrase that we get in a correct answer? Is the phrase being used or mentioned?" It is clear that nothing is settled about *this* question by a choice whether to say, following older usage, that the phrase *gives* the direct object or, following more modern usage, that "direct object" is a name for a sentence-part.

I propose – for a purpose which will appear – to adopt the older usage. Then the question "What is the direct object of the verb in this sentence?" is the same as "What does the sentence say John sent Mary?" and the question "What does the phrase which is the answer to that question communicate to us, i.e. is it being used or mentioned?" can be asked in the form "Is the direct object a bit of language or rather what the bit of language stands for?" – and *this* is now not a mere question of terminology, but a substantive-seeming question of curious perplexity. For someone pondering it may argue as follows: It won't do to say that in this example a book is the direct object. For if we say that we can be asked: "Which book?"; but the sentence isn't being considered as true, and there is no answer to the question "Which book?" except "No book"; and yet without doubt the verb has a direct object, given by the answer "A book". So it must be *wrong*, and not just a matter of terminology, to say that the grammatical phrase "direct object" stands for, not a bit of language, but rather what the bit of language stands for. And, if intentional objects are a sub-class of direct objects, the phrase "intentional object" too will stand for a bit of language rather than what the language stands for; we are evidently not going to have to plunge into the bog made by the fact that in the most important and straightforward sense the phrase giving the intentional object may stand for nothing.

But wait – in that case *must* we not say, "the phrase which *is* the intentional object" rather than "the phrase giving the intentional object"? This is indeed a difficulty. For the intentional object is told in answer to a question "What?" But the answer to "What do they worship?" cannot be that they worship a phrase any more than that they worship an idea. A similar point holds, of course, for direct (and indirect) objects in general.

It may be argued that this is no argument.[2] Perhaps we cannot say "What John is said to have sent is a phrase." But then no more can we say "What John is said to have sent is a direct object" – for the sentence did not say John sent Mary a direct object.

What this shows is that there is a way of taking "The direct object is not a direct object" which makes this true; namely, by assimilating this sentence to "The direct object is not a girl". (One could imagine explaining to a child: "The girl isn't the direct object, but the *book* that John sent.")

Frege's conclusion "The concept horse is not a concept" was based on the same sort of trouble about different uses of expressions. What "*cheval*" stands for is a concept, and what "*cheval*" stands for is a horse; these premisses do not, however, yield the result that if Bucephalus is a horse he is a concept. Similarly, what John is said to have sent Mary is a book, and what John is said to have sent Mary is a direct object; these premisses do not yield the result that if John gave Mary a book, he gave her a direct object.

Frege eventually proposed to deal with the trouble by stipulating that such a phrase as "What 'cheval' stands for" should *only* be used predicatively. A parallel stipulation in our case: "What John is said to have sent Mary is . . ." may only be completed with such expressions as could fill the blank in "John sent Mary . . .".

The stipulation, while harmless, would be based on failure of ear for the different use of the phrase "What John is said to have sent Mary" in the explanation "What John is said to have sent Mary is the direct object of the sentence". But an ear for a different use cannot be dispensed with, as the further course of the argument shows.

The argument began with stating reasons why a direct object can't be

[2] This was argued to me by Mr G. Harman, for which I am obliged to him.

something that the direct-object phrase stands for. Yet one can, one correctly does, say "A book" in answer to the question "What does the sentence 'John sent Mary a book' say John sent Mary?" which asks the same thing as "What is the direct object in that sentence?" Nevertheless the way the phrase "a book" is being used is such that one can't sensibly ask "Which book?"

We must conclude of 'objects' (direct, indirect and likewise intentional) that the object is neither the phrase nor what the phrase stands for. What then is it? The question is based on a mistake, namely that an explanatory answer running say "An intentional (direct, indirect) object is such-and-such" is possible and requisite. But this need not be so. Indeed the only reasonable candidates to be answers are the ones we have failed. But what is the actual use of the term? Given a sentence in which a verb takes an object, one procedure for replying to the question: "What is the object in this sentence?" is to recite the object phrase.

If putting the object phrase in quotes implies that the object – i.e. what John is said to have sent Mary, what the Greeks worshipped – is a piece of language, that is wrong; if its not being in quotes implies that something referred to by the object phrase is the object, that is wrong too. To avoid the latter suggestion one might insist on putting in quotes; to avoid the former one might want to leave them out. One is inclined to invent a special sort of quotes; but the question is how the phrase within such new quotes would function – and if we understand that, we don't need a new sign. So ends the argument.

To repeat, I am not opposing the practice of grammarians and linguists for whom the expression "direct object" is defined as an expression for a phrase; they use that as I use the expression "direct-object phrase". But, as I have argued, the question "What does the sentence say John gave?" is fundamental for understanding either "direct object" or "direct-object phrase" as I am using those expressions; and hence for understanding "direct object" when it is used for a phrase. And though the question is answered (like many questions) by uttering a phrase – in this case "a book" – the phrase has *a special use* in answer to that question "What does the sentence say John gave?" *It* can name neither a piece of language, nor anything that the piece of language names or otherwise relates to, nor indeed anything else. The interest of the question and answer is the rather special interest of getting grammatical understanding. Grammatical understanding and grammatical concepts, even the most familiar ones like sentence, verb, noun, are not so straightforward and down-to-earth a matter of plain physical realities as I believe people sometimes suppose. The concept of a noun, for example, is far less of a physical concept than that of a coin; for someone might be trained to recognize coins with fair success though he knew nothing of money, but no one could be trained to recognize nouns without a great familiarity with language; and yet the concept of a noun is not one which he will automatically have through that familiarity, as he will have that of a coin if he operates with coined money. Indeed the explanations of grammatical terms are only hints at what is really grasped from examples. Thus no one should think that by merely adopting the usage of modern grammarians, for whom the direct object is a word or words, he has avoided handling difficult concepts and remained in a plain man's world of plain things.

> "The direct object is what John sent" (="what the sentence says John sent").
>
> "The intentional object is what *X* was thinking of."

These two sentences are parallel. It is for the sake of parallelism that we opted for the old-fashioned usage of "direct object". For even in that usage, no one will be tempted to think that direct objects as such are a special type of entity. Just this temptation exists very strongly for objects of thought and sensation; that is, for intentional objects, which appear as entities under the names "idea" and "impression".

It may be objected: the context "The sentence says John sent Mary —" is itself intentional. How, then, can my considerations about direct objects throw light on intentional objects? Fully spelled out they are themselves merely examples of sentences whose objects are intentional objects.[3]

The answer is that what is said in the objection is true. But these examples, where we talk about direct objects, are harmless and profitable because certain sorts of suggestion about direct objects are patent nonsense. For example no one would think that if a sentence says John sent Mary a book, what it immediately and directly says he sent her was a direct object, and only in some indirect fashion, via this immediate object, does it say he sent her a book. I want, that is, to use a comparison with patent nonsense about direct objects in order to expose as latent nonsense of just the same kind some very persuasive views about ideas and impressions. Not that ideas and impressions are to be excluded from consideration; but as they enter into epistemology they will be rightly regarded as grammatical notions, whose role is readily misunderstood. And "grammatical" is here being used in its ordinary sense.

We must now ask: does any phrase that gives the direct object of an intentional verb in a sentence necessarily give an intentional object? No. Consider: "These people worship Ombola; that is to say, they worship a mere hunk of wood." (cf. "They worship sticks and stones.") Or "They worship the sun, that is, they worship what is nothing but a great mass of frightfully hot stuff." The worshippers themselves will not acknowledge the descriptions. Their idol is for them a divinized piece of wood, one that is somehow also a god; and similarly for the sun.

An intentional object is given by a word or phrase which gives a *description under which*.

It will help if we consider shooting at, aiming. A man aims at a stag; but the thing he took for a stag was his father, and he shoots his father. A witness

[3] I am indebted for this objection and the discussion of it to Professors Bernard Williams and Arthur Prior and Mr P. T. Geach.

reports: "He aimed at his father." Now this is ambiguous. In the sense in which given the situation as we have described it, this report is true, the phrase "his father" does not give an intentional object. Let us introduce the term "material object": "his father" gives, we shall say, the *material* object of the verb in the sentence "He aimed at his father" in the sense in which this was true. Not because he hit his father – he might after all merely have gone wide of the mark. But because the thing he took for a stag actually was his father. We can ask what he was doing – what he was aiming at – *in that* he was aiming at a stag: this is to ask for another description "*X*" such that in "He was aiming at *X*" we still have an intentional object, but the description "*X*" gives us something that exists in the situation. For example, he was aiming at that dark patch against the foliage. The dark patch against the foliage was in fact his father's hat with his father's head in it.

Thus, the given intentional object (the stag) being non-existent in the situation, we looked for another intentional object until we found one that did exist. Then the phrase giving that intentional object, and any other true description of the existent thing in question, gives the *material* object of "He aimed at . . .".

Does this account depend on the report's being true? No; but if the witness lies or is quite mistaken, all the same he can be questioned about what his report meant. Does he mean the phrase "his father" to give the intentional, or only the material, object? If only the material object, what does he mean by "He aimed at . . ."? That you could see that the man was taking aim, and where his target lay? There might not be true answers to these questions, but the witness has got to pretend there are or be confounded.

And now, for greater ease of expression, I will speak, as is natural, of the material and intentional objects of aiming, of worshipping, of thinking. This should always be interpretable in terms of the verbs and their objects.

There need not be a material object of aiming. If a man were totally hallucinated, and, shooting at something in his hallucinatory scene, hit his father, that would not make his father the *material* object of his aiming. Similarly, if there is no description, still giving the intentional object of worship, which describes anything actual, the worshippers, materially speaking, worship a nothing, something that does not exist.

Not that it will then do to say "They worship nothing", but only: "What they worship is nothing." For "They worship nothing" would imply that no sentence "They worship such-and-such" will be true; and in the case supposed some such sentence is true.

Questions about the identity of an intentional object, when this cannot be reduced to the identity of a *material* object, are obviously of some interest. How do we decide that two people or peoples worship or do not worship the same god? Again, when a proper name is obscure and remote in its historical reference, like "Arthur", the question may arise whether two people are thinking of the same man – if they have different, incompatible, pictures of him.

But I perceive that my saying "when this cannot be reduced to the identity of a *material* object" may mislead: for by *material* objects I do not mean what are now called "material objects" – tables, planets, lumps of butter and so on. To give a clear instance: a debt of five dollars is not a material object in this latter sense; but given that someone had contracted such a debt, my thought "that debt of five dollars" would have as its material object something described and indicated by the phrase giving the intentional object of my thought. When it is beyond question that the phrase giving an intentional object does describe and indicate a material object in this sense, then the question as to the identity of the intentional object reduces to the question as to the identity of the material object. Are we referring to the same debt? That is, perhaps, not too difficult to establish. But when either there is no real debt or it is very obscure whether there is, the case is altered.

The fact that we can use the concept of identity in connection with intentional objects should not lead us to think there is any sense in questions as to the kind of existence – the ontological status – of intentional objects as such. All such questions are nonsensical. Once more we can clear our heads by thinking of direct objects. The answer to "What is the direct object in 'John sent Mary a book'?" is "A book". This is the right answer as much when the sentence is false as when it is true, and also when it is only made up, as it is in this case, to illustrate a point. It is evident nonsense to ask about the mode of existence or ontological status of the direct object as such: or to ask what kind of thing *a book* is, as it is thought of in answer to the question about the direct object.

II *Sensation*

In the philosophy of sense-perception there are two opposing positions. One says that what we are immediately aware of in sensation is sense-impressions, called "ideas" by Berkeley and "sense-data" by Russell. The other, taken up nowadays by "ordinary language" philosophy, says that on the contrary we at any rate *see* objects (in the *wide* modern sense which would include, e.g. shadows) without any such intermediaries. It is usually part of this position to insist that I can't see (or, perhaps, feel, hear, taste or smell) something that is not there, any more than I can hit something that is not there: I can only *think* I see (etc.) something if it isn't there, or only in some extended usage of "see" do I see what isn't there. I shall say most about seeing, as most people do in discussing this topic. The other verbs are for good reasons (which aren't very relevant to my topic) often treated rather differently, especially by ordinary language philosophy.

I wish to say that both these positions are wrong; that both misunderstand verbs of sense-perception, because these verbs are intentional or essentially have an intentional aspect. The first position misconstrues intentional objects as material objects of sensation; the other allows only *material* objects of sensation; or at any rate does not allow for a description of what is seen

which is e.g. neutral as between its being a real spot (a stain) or an after-image, giving only the content of an experience of seeing concerning which one does not yet know whether one is seeing a real spot or an after-image.[4]

To see the intentionality of sensation it is only necessary to look at a few examples which bring it out.

(1) "When you screw up your eyes looking at a light, you see rays shooting out from it."
(2) "I see the print very blurred: is it blurred, or is it my eyes?"
(3) "Move these handles until you see the bird in the nest." (Squint-testing apparatus; the bird and the nest are on separate cards.)
(4) "I see six buttons on that man's coat, I merely see a lot of snow flakes framed by this window-frame – no definite number."
(5) ". . . a mirage. An approaching pedestrian may have no feet (they are replaced by a bit of sky)."[5]
(6) "With this hearing aid, when you talk I hear some screeching noises; no low tones and the consonants are very indistinct."
(7) "I hear a ringing in my ears."
(8) "I heard a tremendous roaring noise outside, and wondered with alarm for a moment what great machine or floodwater could be making it. And then I realized that it was only my little dog snoring close at hand."[6]
(9) "Do you know how a taste can sometimes be quite indeterminate until you know what you are eating?"
(10) "I keep on smelling the smell of burning rubber when, as I find out, there is no such thing."

Someone who wishes to say that the verbs of sense are used right in normal cases *only* with real things as objects, and even with real things correctly characterized, may say that these are exceptional uses. Either the context (eye-testing apparatus) or what is said, with the tone of voice and special emphasis appropriate to it, shows this. There was presumably a definite number of snowflakes falling so as to be seen from a certain position, and that was the number seen; only the subject did not know how many there were, was not able to tell by looking as he could tell the number of buttons on the coat. He expressed this by saying he did not *see* a definite number of snowflakes; but this is an odd use of "see", different from the more normal use we get in the following example:

(11) "I saw someone in the study just now." "Nonsense! You can't have, because there isn't anyone there." "Well, I wonder what I saw, then."

Now this may be; on the other hand the oculist testing the degree of a

[4] I am obliged to Professor Frank Ebersole for telling me of an experience of his which supplied this example.
[5] Example from M. Luckiesh.
[6] Example from W. James.

squint does not have to teach a new use of "see" or of "I see a (picture of a) bird in a nest" before he can ask "Do you see the bird in the nest?" – the bird-picture and the nest-picture being in fact spatially separated. To call such a use "new" simply means that some difference between it and what is being called the old use strikes us as important.

There is indeed an important difference; though it is wrong to regard the uses which it marks as, so to speak, *deviant*, for our concepts of sensation are built up by our having *all* these uses. The difference we are attending to is that in these cases, object phrases are used giving objects which are, wholly or in part, merely intentional. This comes out in two features: neither possible non-existence (in the situation), nor indeterminacy, of the object is any objection to the truth of what is said.

Now 'ordinary language' views and 'sense-datum' views make the same mistake, that of failing to recognize the intentionality of sensation, though they take opposite positions in consequence. This failure comes out clearly on the part of an ordinary-language philosopher if he insists that what I say I see must really be there if I am not lying, mistaken, or using language in a "queer", extended (and therefore discountable) way.

The Berkeleyan sense-datum philosopher makes the same mistake in his insistence that, e.g., one sees visual impressions, visual data. I would say that such a philosopher makes an incorrect inference from the truth of the grammatical statement that the intentional object, the impression, the visual object, is what you see. He takes the expression "what you see" materially. "The visual impression is what you see", which is a proposition like "The direct object is what he sent", is misconstrued so as to lead to "You see an impression", as the other never would be misconstrued so as to lead to "He sent her a direct object".

This is a more interesting and permanently tempting mistake than the other, whose appeal is merely that of a common-sense revolt against a Berkeleyan type of view. But both doctrines have a great deal of point. To take the 'ordinary language' doctrine:

First, what I shall call the material use of verbs of sense exists. The material use of "see" is a use which demands a *material* object of the verb. "You can't have seen a unicorn, unicorns don't exist." "You can't have seen a lion, there wasn't any lion there to see." These uses are quite commonplace. It is not merely that the object-phrase is taken materially – as we have seen, that may be the case with an intentional verb without reflecting on its intentionality. Here the verb "to see" is not allowed to take a *merely* intentional object; non-existence of the object (absolutely, or in the situation) is an objection to the truth of the sentence. We see the double use of the verb "see" by contrasting it with "worship". No one would ever say: "They cannot have worshipped unicorns, because there are no such things."

Second, the words giving the object of a verb of sense are necessarily most often intended as giving *material* objects of sense: for this is their primary application. To see this, consider the following. Suppose a bright red plastic

toy elephant looks greyish-brown to me in a certain light. Only if I do not know that the greyish-brown colour is mere appearance do I say without any special context (e.g. that of describing impressions), or apology, or humour: "I see a greyish-brown plastic toy elephant." This is because we understand the description-of-an-appearance "greyish-brown" by understanding the description "greyish-brown": this describes what the appearance is of. To do that, it must in the first instance be a description of such a thing as it would be true of (for the appearance is an appearance of that) – really, and not merely in appearance: this will be its primary application. But, being a description of a sensible property, it must also in its primary application enter into the object phrases for the appropriate verbs of sense, since we get to know sensible properties by the appropriate senses.

Further, we ought to say, not: "Being red is looking red in normal light to the normal-sighted," but rather "Looking red is looking as a thing that *is* red looks in normal light to the normal-sighted." For if we ought rather to say the first, then how do we understand "looking red"? Not by understanding "red" and "looking". It would have to be explained as a simple idea; and so would looking any other colour. It may be replied: These all are simple ideas; "looking yellow" and "looking red" are the *right* expressions for what you show someone when you show him yellow and red, for he will only learn "yellow" and "red" from the examples if they look yellow and look red; so it is *looking-yellow* and *looking-red* that he really gets hold of and has been introduced to, even though you *say* you are explaining "yellow" and "red". This would come to saying that in strictness "looking" should be part of every colour word in reports of perception: it will then cease to perform the actual function of the word "looking". It was plausible to say: Only if it looks red to him will he learn what is meant; but wrong to infer: What he then grasps as the correlate of the word "red" is a red look. Even granted that he knows he is to learn the name of a colour, still it invites misunderstanding to rely on something that only *looks* red to teach him the word; if he notices that it only looks red, how natural for him to suppose that "red" was the name of the colour that it actually *is*. If you tell him: "It's the colour that this 'looks'," this presupposes that "looks *C*" and "*C*" are originally, and not just subsequently, distinct: that, in short, "being red" is not after all to be explained as a certain looking-red.

Again, things do not always look the same shape, colour, size and so on, but we commonly look at and describe them, saying, e.g., "It's rectangular, black and about six foot in height," without paying attention to how they look – indeed we might say that often things *look* to us, strike us, not as they look but as they are! (Conviction that *only* so is "looks" used rightly was the cause of confusion to an over-confident ordinary-language philosopher on an occasion famous in Oxford: F. Cioffi brought in a glass vessel of water with a stick in it. "Do you mean to say", he asked, "that this stick does not look bent?" "No," said the other bravely: "It looks like a straight stick in water." So Cioffi took it out and it *was* bent.)

So much at least there is to be said on the side of the "ordinary-language" philosopher. But, turning to the sense-impression philosophy, how much it points out and can investigate which often gets querulously dismissed by the other side! There is such a thing as simply describing impressions, simply describing the sensible appearances that present themselves to one situated thus and thus – or to *myself*.

Second, the sense-impression philosophy will be right in its way of taking the Platonic dictum: "He who sees must see something." Plato compared this to "He who thinks must think something," and has sometimes been criticized on the ground that "seeing" is a relation of a subject to an object in the modern sense of that last word, while thinking is different: that such-and-such is the case isn't a thing. But "He who sees must see something" is being wrongly taken if taken as meaning: "Whenever anyone can rightly be said to see, there must be something there, which is what he sees." Taken in that sense, it is not true; to say it is true is to legislate against all except the material use of "see". The sense in which it is true is that if someone is seeing, there is some content of his visual experience. If he says he can see ("can see" is English idiom for "is seeing") we can ask him "What can you see?" He may say "I don't know". Perhaps that means that he doesn't know what the material object of his seeing is; perhaps simply that he is at a loss to make out *what* what he (in any sense) sees *looks like*. But then we can say: well, at any rate, describe what colours, what variation of light and dark you see. He may say: "It's frightfully difficult, it all changes so fast, so many colours shifting all the time, I can't describe it, it doesn't stay long enough" – and that's a description. But he cannot say: "How do you mean, what I see? I only said I could see, I didn't say I could see something – there's no need of a '*what*' that I see." That would be unintelligible.

This brings out the third point in favour of the sense-impression philosophy, which offers it some support even in its strict Berkeleyan form. The minimum description that must be possible if someone can see, will be of colours with their variations of light and darkness. One cannot say "Colour, light and dark? No question of any such things," in response to a *present* enquiry about what one sees.

That is to say, it is so with us. Perhaps we could imagine people whose language has no colour vocabulary, though they are sighted, i.e. they use eyes and need light to get about successfully, etc. A man of such a people, taught to read by sight, learns names of letters, could read out words which were black on white, but could not understand the words "black" and "white". We'd say we do not know 'how he tells' the words, the shapes. But is that to say anything but that for us appeal to colours is used in an account of how we tell shapes? Whereas perhaps for him there is in this sense no such thing as a 'how he tells' – any more than there is for us with the colours themselves. We don't ask for a 'how we tell' it's red, as we ask for a 'how we tell' it's the word "red" and accept as part of the answer "by seeing these shapes, i.e. colour patches of these shapes". We may wonder "How could

there be such recognition of a thing like the pattern of a word – *unmediated* recognition? How could it but be mediated by perception of colour?" (One of the origins of the notion of simple ideas, elements.) But although in this case we have an account of the perception of the pattern as mediated by the perception of colour, think of our recognition of human expressions. We feel that this is the *kind* of thing to be mediated, but fail in our attempts to describe the elements and their arrangements, seeing which we recognize a cheerful or ironical expression. But, one may say, optically speaking he must be being affected by light of the wavelengths belonging to the different colours. Yes – but does that show that, so to speak, the content of a colour concept is pushed into him, so that all he has to do is utter it in a name, whose use he will later make to fit with other people's in its range of application? I believe this is thought. (cf. Quine about "square" and each man's retinal projection of a square tile.)[7] Formulated, this loses its plausibility. For one thing, the optical process does not exhibit anything to the man in whom it takes place. For another, no concept is simply given; every one involves a complicated technique of application of the word for it, which could not just be presented by an experience-content. The fact that there is no 'how we tell' about colour-recognition does not mean that training in practices – most strikingly the practices comprising that technique of application – is not as necessary for the acquisition of colour concepts as those of substances or square roots.

Pursuant to this false conception of the primitively given, Berkeley – and Russell – thought that all else in description of the seen, all besides the arrangement of colour patches in the visual field, was inference and construction. This is not acceptable. There are impressions of distance and size, for example, independent of assumptions about what a thing is. One may be utterly perplexed what a thing is just because one is seeing it as at a different distance from the right one, and hence as the wrong size. Or vice versa. I once opened my eyes and saw the black striking surface of a matchbox which was standing on one end; the other sides of the box were not visible. This was a few inches from my eye and I gazed at it in astonishment wondering what it could be. Asked to describe the impression as I remember it, I say: "Something black and rectangular, on end, some feet away, and some feet high." I took it for three or four feet distant, and it looked, if anything, like a thick post, but I knew there could be no such thing in my bedroom. Or I have taken a small black prayer book for a great family Bible sort of volume, judging that it lay on a footrest some feet away instead of a nearby ledge nearer eye-level. These were not judgements of distance based on identifications of things – the supposition of what thing it might be was based on an impression of size which went with a false impression of distance.

Departing, then, from Berkeley, we can note that descriptions of visual impressions can be very rich and various. There can be impressions of depth and distance and relative positions and size; of kinds of things and kinds of

[7] *Word and Object* (Cambridge, Mass., 1960), p. 7.

stuff and texture and even temperature; of facial expression and emotion and mood and thought and character; of action and movement (in the *stationary* impression) and life and death. Even within the compass of the description "colours with their variations of light and shade" there are diverse kinds of impression.

It remains to sort out the relations between the intentional and material objects of sensation; as I have done most of the time, I will concentrate on seeing.

While there must be an intentional object of seeing, there need not always be a material object. That is to say "*X* saw *A*" where "saw" is used materially, implies some proposition "*X* saw —" where "saw" is used intentionally; but the converse does not hold. This leads to the feeling that the intentional use is somehow prior to the material use. The feeling seems to run contrary to the recognition, the feeling, that for descriptions of objects of sight the material application is the prior one. Both feelings are – legitimately – satisfied by allowing that an intentional object is necessarily involved in seeing, while granting that this does not confer epistemological priority on purely intentional sentences, which indeed, in a host of the most ordinary cases of reported seeing, are never formulated or considered.

John Austin, who opposed the view that there are two senses of "see" according as the seeing has to be veridical or not, remarked casually that there were perhaps two senses of "object of sight". I think it was in this connection that he contrasted "Today I saw a man born in Jerusalem" and "Today I saw a man shaved in Oxford" – both said in Oxford. At any rate, one says, you didn't *see* him born today; perhaps you did see someone being shaved. So the one description, while true of what you saw, in a sense does not give what you saw. A description which is true of a material object of the verb "to see", but which states something that absolutely or in the circumstances "you can't have *seen*", necessarily gives *only* a material object of seeing.

In speaking of the material object of aiming, I said that if a man aimed at that dark patch against the foliage, and that patch was his father's hat with his father's head in it, then his father was a material object of his aim; but if he aimed at some patch in a totally hallucinatory scene, and hit his father, you could not say that.

Now if we try to apply this explanation to the case of seeing we run into difficulties which reflect back on the case of aiming. But in the case considered the material object of aiming was arguably an *intentional* object of seeing. For what else – it might be asked – is a dark patch against the foliage?

This may seem to plunge us into confusion. For surely what is *only* an intentional object of seeing can't be a material object of aiming? Then when does a description give a material object of sight? One kind of case we have seen: when a description is true of what is seen, but does *not* give an intentional object. "I see a man whose great uncle died in a lunatic asylum" – the relative clause gives an absolutely non-intentional description. "I see a girl who has a mole between her shoulder-blades" – in the circumstances it gives

a non-intentional description. For she is facing me, etc. "You can't have *seen* that," one says.

But why? If I can't see that, why can I see Professor Price's tomato? It has a backside that I don't see. Mr Thompson Clarke draws our attention to the fact that a view of a tomato and a half-tomato may be exactly the same. That is so; but it is not like the fact that a view of someone with and without a mole between his shoulder blades may be exactly the same. If you look at a tomato and take only a single view, you *must* see what *might* be only a half tomato: that is what seeing a tomato is. Whereas there is a view of the mole; and no front view *is* a view of a mole between the shoulder blades. Such a mole does not stamp the front view as may approaching death or a load of troubles, and so there is no impression of it – just as there is no "born-in-Jerusalem" look about a man.

But a material object of seeing is not necessarily given by a description of what is before my eyes when they are open and I am seeing; if I am totally hallucinated, then in no sense do I see what is before my eyes. Thus it is essential to a material object of seeing that it is given by a description which is true of *what is seen*; and we have to enquire into the significance here of this phrase "*what is seen*".

The problem is this: there is a material object of φ-ing if there is a phrase giving an intentional object of φ-ing which is also a description of what exists in a suitable relation to the φ-er, Now this can't be a description of what exists merely by describing the intentional object of some *other* act (he aims at the dark patch that he sees); if simply describing an intentional object of φ-ing will not – as of course it will not – guarantee that we have described a material object of φ-ing, then how can it give a material object of some other verb, ψ-ing?

All would be plain sailing if we could say: we have a material object of sight only if *some* intentional description is also true of what really – physically – exists. And perhaps we can say that the dark patch against the foliage is not merely an intentional object of seeing; there really is a dark object or a region of darkness there.

But this is not always the case when we see. Suppose I have defective sight: all I see is a shiny blur over there. That blur, we say, is my watch. We therefore say I see my watch, though very indistinctly; and I want to say that my watch is the material object of seeing. But I may not be able to see it as a watch; all I see is a shiny blur. But the description "a shiny blur" is not true of anything that phsyically exists in the context. Supposing the father had a dark hat on, it would follow that, to mention the puzzle that perplexed Moore for so long, the dark patch against the foliage was *part of the surface of a material object (modern sense)*; but certainly 'a blur' is no part of the surface of my watch. But it may be I have no other description of what I see than "a shiny blur over there". So is there any intentional description which is also a description of a material object of sight?

Yes; for even if my watch is not a blur, it is a shiny thing and it is over there.

Suppose I had said: I see a roughly triangular red blur here, and some causal connection via the visual centres in the brain could have been discovered between that and the presence of my watch over there – would it have been right to say: "What I am seeing is my watch"? I believe not.

An interesting case is that of *muscae volitantes*, as they are called. You go to the doctor and you say: "I wonder if there is something wrong with my eyes or my brain? I see" – or perhaps you say "I *seem* to see" – "floating specks before my eyes." The doctor says: "That's not very serious. They're there all right" (or: "You see them all right") – "they are just the floating debris in the fluids of the eye. You are a bit tired and so your brain doesn't knock them out, that's all." The things he says you see are not *out there* where you say you see them – *that* part of your intentional description is not true of anything relevant; but he does not say that what you are seeing is that debris *only* because the debris is the cause. There really are floating specks. If they caused you to see a little red devil or figure of eight, we should not say you saw them. It may be possible to think of cases where there is nothing in the intentional object that suggests a description of what is materially being seen. I doubt whether this could be so except in cases of very confused perception – how could a very definite intentional description be connected with a quite different material object of seeing? In such cases, if we are in doubt, we resort to moving the supposed material object to see if the blurred, not colour-true and misplaced image of it moves.

When you said: "I see" – believing that the objects were quite illusory – you *intended* your description purely as an intentional one; you were giving the words "floating specks" a secondary application. It came as a surprise to you that you would have had the right to intend the words materially. In the well-known case of H. H. Price's mescaline illusion, when without any derangement of his judgement he was able to describe what he saw – a great pile of leaves on his counterpane, which he knew not to be there – we again have a secondary application: the words "a pile of leaves" were intended *only* as a description of an impression.

It is important to notice that very often there is no answer to the question whether people intend the word "see" in its *material* use or not: that is, whether they are so using the word "see" that they would have to take it back supposing that what they said they saw was not there. If they were mis-seeing something that was there, they would usually want to correct themselves, finding out 'what they really saw'. But what if the seeing were hallucinatory?

The question would be: supposing that turned out to be the case, would you claim that you mean "see" in such a way that all you have to do is alter your intentions for the description of the object, from intending it in its *primary* application as a description of the *material* object of sight to intending it in a *secondary* application as a description of a mere *impression*?

Faced with such a question, we have in general the right to reject it, saying like Tommy Traddles: But it isn't so, you know, so we won't suppose it if you don't mind. And even if we have not this right, we generally entertain no

such supposition and *therefore* are unprepared with an answer. We need not have determinately meant the word "see" one way or the other.

We may make a similar point about 'phantom limb'. I take the part of the body where pain is felt to be the object of a transitive verb-like expression "to feel pain in —". Then when there is, e.g., no foot, but *X*, not knowing this, says he feels pain in his foot, he may say he was wrong ("I did not see a lion there, for there was no lion") or he may alter his understanding of the phrase "my foot" so that it becomes a purely intentional object of the verb-like expression. But it need not be determined in advance, in the normal case of feeling pain, whether one so intends the expression "I feel pain in —" as to withdraw it, or merely alters one's intentions for the description of the place of the pain, if one should learn that the place was missing.

2 The First Person

Descartes and St Augustine share not only the argument *Cogito ergo sum* – in Augustine *Si fallor, sum* (*De Civitate Dei*, XI, 26) – but also the corollary argument claiming to prove that *the mind* (Augustine) or, as Descartes puts it, *this I*, is not any kind of body. "I could suppose I had no body," wrote Descartes, "but not that I was not", and inferred that "this I" is not a body. Augustine says "The mind knows itself to think", and "it knows its own substance": hence "it is certain of being that alone, which alone it is certain of being" (*De Trinitate*, Book X). Augustine is not here explicitly offering an argument in the first person, as Descartes is. The first-person character of Descartes' argument means that each person must administer it to himself in the first person; and the assent to St Augustine's various propositions will equally be made, if at all, by appropriating them in the first person. In these writers there is the assumption that when one says "I" or "the mind", one is naming something such that the knowledge of its existence, which is a knowledge of itself as thinking in all the various modes, determines what it is that is known to exist.

Saul Kripke has tried to reinstate Descartes' argument for his dualism. But he neglects its essentially first-person character, making it an argument about the non-identity of *Descartes* with his own body. Whatever else is said, it seems clear that the argument in Descartes depends on results of applying the method of doubt.[1] But by that method Descartes must have doubted the existence of the man Descartes: at any rate of that figure in the world of his time, that Frenchman, born of such-and-such a stock and christened René;

[1] *Principles of Philosophy*, I, LX, contains Descartes' best statement, which is I think immune to the usual accusation of substitutional fallacy: "Each of us conceives of himself as a conscious being, and can in thought exclude from himself any other substance, whether conscious or extended; so from this mere fact it is certain that each of us, so regarded, is really distinct from every other conscious substance and from every corporeal substance. And even if we supposed that God had conjoined some corporeal substance to such a conscious substance so closely that they could not be more closely joined, and had thus compounded a unity out of the two, yet even so they remain really distinct" (*Philosophical Writings*, trans. G. E. M. Anscombe and P. T. Geach). Rendering Descartes' premise here as "I can conceive myself not to include or be my body", we come close to Kripke's version (but in the first person) "Possibly I am not *A*", where "*A*" means my body. But why can I so conceive myself if not because I can doubt the existence of my body?

But "doubting" here does not mean merely reflecting that I am ignorant of the existence of my body though not of myself. So understood, the argument would indeed involve the substitutional fallacy. "Doubting" means clearly understanding that the existence of my body is not guaranteed by something which is throughly understood, and is all I am sure of: the existence of myself. We see the importance of the premise supplied by St Augustine "The mind knows its own substance".

From Samuel Guttenplan (ed.), *Mind and Language: Wolfson College Lectures 1974* (Oxford, 1975).

but also, even of the man – unless a man isn't a sort of animal. *If*, then, the non-identity of himself with his own body follows from his starting-points, so equally does the non-identity of himself with the man Descartes. "I am not Descartes" was just as sound a conclusion for him to draw as "I am not a body". To cast the argument in the third person, replacing "I" by "Descartes", is to miss this. Descartes would have accepted the conclusion. That mundane, practical, everyday sense in which it would have been correct for him to say "I am Descartes" was of no relevance to him in these arguments. That which is named by "I" – *that*, in *his* book, was not *Descartes*.

It may seem strange to say: "The non-identity of himself with Descartes was as valid a conclusion as the other" and not treat this as already a *reductio ad absurdum*. For is that phrase not equivalent to "the non-identity of *Descartes* with Descartes"?

No. It is not. For what is in question is not the ordinary reflexive pronoun, but a peculiar reflexive, which has to be explained in terms of "I". It is the reflexive called by grammarians the 'indirect reflexive' and there are languages (Greek, for example) in which there is a special form for it.[2]

"When John Smith spoke of James Robinson he was speaking of his brother, but he did not know this." That's a possible situation. So similarly is "When John Smith spoke of John Horatio Auberon Smith (named in a will perhaps) he was speaking of himself, but he did not know this." If so, then 'speaking of' or 'referring to' oneself is compatible with not knowing that the object one speaks of is oneself.

Yet we are inclined to think that "It's the word each one uses in speaking of himself" explains what "I" names, or explains "I" as a 'referring expression'. It cannot do so if "He speaks of himself" is compatible with ignorance and we are using the reflexive pronoun, in both cases, in the ordinary way.

Nor can we explain the matter, as we might suppose, by saying "'I' is the word each one uses when he knowingly and intentionally speaks of himself". For did not Smith knowingly and intentionally speak of Smith? Was not the person he intended to speak of – Smith? and so *was* not the person he intended to speak of – himself?

It may be said: "Not in the relevant sense. We all know you can't substitute every designation of the object he intended to speak of and keep the statement about his intention true." But that is not the answer unless the reflexive pronoun itself is a sufficient indication of the way the object is specified. And that is something the ordinary reflexive pronoun cannot be. Consider: "Smith realizes (fails to realize) the identity of an object he calls 'Smith' with himself." If the reflexive pronoun there is the ordinary one, then it specifies

[2] *ἑ, οὗ, οἷ*. See Thucydides II. 13. The form is rare. Credit for discerning the indirect reflexive in English, which does not have a distinct form for it, belongs in the present day to H.-N. Castaneda in "The Logic of Self-Knowledge", *Noûs*, I (1967), 9–22. But his presentation is excessively complicated and I believe it has not attracted enough attention to the substantive point.

for us who frame or hear the sentence an object, whose identity with the object he calls "Smith" Smith does or doesn't realize: namely the object designated by our subject word "Smith". But that does not tell us what identity Smith himself realizes (or fails to realize). For, as Frege held, there is no path back from reference to sense; any object has many ways of being specified, and in this case, through the peculiarity of the construction, we have succeeded in specifying an object (by means of the subject of our sentence) without specifying any conception under which *Smith*'s mind is supposed to latch on to it. For we don't want to say "Smith does not realize the identity of Smith with Smith".

We only have to admit a failure of specification of the intended identity, if we persist in treating the reflexive in "He doesn't realize the identity with himself" as the ordinary reflexive. In practice we have no difficulty at all. We know what we mean Smith doesn't realize. It is: "I am Smith." But if that is how we understand that reflexive, it is not the ordinary one. It is a special one which can be explained only in terms of the first person.

If that is right, the explanation of the word "I" as 'the word which each of us uses to speak of himself' is hardly an explanation! – At least, it is no explanation if that reflexive has in turn to be explained in terms of "I"; and if it is the ordinary reflexive, we are back at square one. We seem to need a sense to be specified for this quasi-name "I". To repeat the Frege point: we haven't got this sense just by being told which object a man will be speaking of, whether he knows it or not, when he says "I". Of course that phrase "whether he knows it or not" seems highly absurd. His use of "I" surely guarantees that he does know it! But we have a right to ask *what* he knows; 'if "I" expresses a way its object is reached by him, what Frege called an "Art des Gegebenseins", we want to know what that way is and how it comes about that the only object reached in that way by anyone is identical with himself.

To say all this is to treat "I" as a sort of proper name. That's what gets us into this jam. Certainly "I" functions syntactically like a name. However, it has been observed not to be a proper name. Now this observation may strike us as obvious enough in a trivial sense. After all, we don't call it a proper noun but a personal *pro*noun. It is at any rate not an ordinary proper name. It could not have a lot of the characteristic use of a proper name. For if it is such, it is one that everyone has, and, worse still, one that each person uses only to refer to that person that he himself is. So it's no use for introducing people to one another, or for calling to someone, or for summoning him. And while it might be used as a signature (like the signature of an aged and doddering parson that I heard of, on someone's marriage lines: Me, Vicar), one would be quite dependent on other clues to the identity of the signatory. If this were the only name anyone had, the situation would be worse than it is for a bank in a Welsh village. These inconveniences are avoided, of course, because there are other more various proper names which people have as well. So the observation that "I" is not a proper name seems to reduce to the

triviality that we perhaps would not *call* a word a proper name if everyone had it and used it only to speak of himself. – But is even that true? After all, all Sikhs seem to be called "Singh". So the real difference lies in that one point that each one uses the name "I" only to speak of himself. Is that a ground not to call it a proper name? Certainly to the eyes of our logicians it is a proper name. Are their eyes dim? Or is it really logically a proper name?

Let us ask: is it really true that "I" is only not called a proper name because everyone uses it only to refer to himself? Let us construct a clear case of just such a name. Imagine a society in which everyone is labelled with two names. One appears on their backs and at the top of their chests, and these names, which their bearers cannot see, are various: "*B*" to "*Z*" let us say. The other, "*A*", is stamped on the inside of their wrists, and is the same for everyone. In making reports on people's actions everyone uses the names on their chests or backs if he can see these names or is used to seeing them. Everyone also learns to respond to utterance of the name on his own chest and back in the sort of way and circumstances in which we tend to respond to utterance of our names. Reports on one's own actions, which one gives straight off from observation, are made using the name on the wrist. Such reports are made, not on the basis of observation alone, but also on that of inference and testimony or other information. *B*, for example, derives conclusions expressed by sentences with "*A*" as subject, from other people's statements using "*B*" as subject.

It may be asked: what is meant by "reports on one's own actions"? Let us lay it down that this means, for example, reports issuing from the mouth of *B* on the actions of *B*. That is to say: reports from the mouth of *B* saying that *A* did such-and-such are prima facie verified by ascertaining that *B* did it and are decisively falsified by finding that he did not.

Thus for each person there is one person of whom he has characteristically limited and also characteristically privileged views: except in mirrors he never sees the whole person, and can only get rather special views of what he does see. Some of these are specially good, others specially bad. Of course, a man *B* may sometimes make a mistake through seeing the name "*A*" on the wrist of another, and not realizing it is the wrist of a man whose other name is after all not inaccessible to *B* in the special way in which is own name ("*B*") is.

(It may help some people's imagination if we change the example: instead of these rather inhuman people, we suppose machines that are equipped with scanning devices are marked with signs in the same way as the people in my story were marked with their names, and are programmed to translate what appears on the screens of their scanners into reports.)

In my story we have a specification of a sign as a name, the same for everyone, but used by each only to speak of himself. How does it compare with "I"? – The first thing to note is that our description does not include self-consciousness on the part of the people who use the name "*A*" as I have described it. They perhaps have no self-consciousness, though each one knows a lot about the object that he (in fact) is; and has a name, the same as

everyone else has, which he uses in reports about the object that he (in fact) is.

This – that they have not self-consciousness – may, just for that reason, seem not to be true. *B* is conscious of, that is to say he observes, some of *B*'s activities, that is to say his own. He uses the name "*A*", as does everyone else, to refer to himself. So he is conscious of himself. So he has self-consciousness.

But when we speak of self-consciousness we don't mean that. We mean something manifested by the use of "I" as opposed to "*A*".

Hence we must get to understand self-consciousness. Unsurprisingly, the term dates only from the seventeenth century and derives from philosophy. Getting into ordinary language, it alters, and by the nineteenth century acquires a sense which is pretty irrelevant to the philosophical notion: it comes to mean awkwardness from being troubled by the feeling of being an object of observation by other people. Such a change often happens to philosophical terms. But this one also gets into psychology and psychiatry, and here its sense is not so far removed from the philosophical one.

The first explanation of self-consciousness that may occur to someone, and what the form of the expression suggests, is this: it is consciousness of a self. A self will be something that some things either have or are. If a thing has it it is something connected with the thing, in virtue of which the thing that has it is able to say, and mean, "I". It is what he calls "I". Being able to mean "I" is thus explained as having the right sort of thing to call "I". The fanciful use of the word, if someone should put a placard "I am only a waxwork" on a wax policeman, or in the label on the bottle in *Alice in Wonderland* "Drink me", is a pretence that the objects in question have (or are) selves. *The self* is not a Cartesian idea, but it may be tacked on to Cartesian Ego theory and is a more consequent development of it than Descartes' identification of 'this I' with his soul. If things are, rather than having, selves, then a self is something, for example a human being, in a special aspect, an aspect which he has as soon as he becomes a 'person'. "I" will then be the name used by each one only for himself (this is a direct reflexive) and precisely in that aspect.

On these views one would explain "self" in "self-consciousness" either by explaining what sort of object that accompanying self was, or by explaining what the aspect was. Given such explanation, one might have that special 'way of being given' of an object which is associated with the name one uses in speaking of it.

Now all this is strictly nonsensical. It is blown up out of a misconstrue of the reflexive pronoun. That it is nonsense comes out also in the following fact: it would be a question what guaranteed that one got hold of the right self, that is, that the self a man called "I" was always connected with him or was always the man himself. Alternatively, if one said that "the self connected with a man" meant just the one he meant by "I" at any time, whatever self that was, it would be by a mere favour of fate that it had anything else to do with him.

But "self-consciousness" is not any such nonsense. It is something real, though as yet unexplained, which "I"-users have and which would be lacking to "*A*"-users, if their use of "*A*" was an adequate tool for their consciousness of themselves.

The expression "self-consciousness" can be respectably explained as 'consciousness that such-and-such holds of oneself'. Nor should we allow an argument running: since the occurrence of "oneself" is just like the occurrence of "himself" which left us perfectly well understanding what Smith failed to realize, the word "self" must itself connote the desired 'way of being given' that is associated with "I" as (logically speaking) a proper name. We must reject this argument because "oneself" is here nothing but the indirect reflexive: that is to say, the reflexive of indirect speech. Understanding indirect speech we know what the related direct speech is. That is all.

These considerations will lack appeal. The question was, what does "I" stand for? If that question is asked, and "I" is supposed to stand for its object as a proper name does, we need an account of a certain kind. The use of a name for an object is connected with a conception of that object. And so we are driven to look for something that, for each "I"-user, will be the conception related to the supposed name "I", as the conception of a city is to the names "London" and "Chicago", that of a river to "Thames" and "Nile", that of a man to "John" and "Pat". Such a conception is requisite if "I" is a name, and there is no conception that can claim to do the job except one suggested by 'self-consciousness'. That is why some philosophers have elaborated the notion of 'selves' (or 'persons' defined in terms of self-consciousness) and conducted investigations to see what such things may be. And just as we must be continuing our reference to the same city if we continue to use "London" with the same reference, so we must each of us be continuing our reference to the same self (or 'person') if we continue to use "I" with the same reference.

This led to an imaginative *tour de force* on the part of Locke: might not the thinking substance which thought the thought "I did it" – the genuine thought of agent-memory – nevertheless be a different thinking substance from the one that could have had the thought: "I am doing it" when the act was done? Thus he detached the identity of the self or 'person' from the identity even of the thinking being which does the actual thinking of the I-thoughts.

Considerations about reflexive pronouns are certainly not going to dam up the flood of inquiries about 'the self' or 'selves', so long as "I" is treated as a name and a correlative term is needed for its type of object. Nevertheless, these are embarrassing credentials for such inquiries. And a self *can* be thought of as what "I" stands for, or indicates, without taking "I" as a proper name. The reasons for considering it as a proper name were two: first, that to the logician's eye it is one, and second, that it seemed to be just like our "*A*" (which was clearly a proper name) except that it expressed 'self-consciousness'. So we tried to explain it as a proper name of a self. Now a lot

of people who will have no objection to the talk of 'selves' will yet feel uneasy about calling "I" a proper name of a self or anything else. I assume it was made clear that the different reference in each mouth was not an objection (there is no objection to calling "*A*" a proper name), and so there is some other reason. The reason, I think, is that, so understood, a repeated use of "I" in connection with the same self would have to involve a reidentification of that self. For it is presumably always a use in the presence of its object! There is no objection to the topic of reidentification of selves – it is one of the main interests of the philosophers who write about selves – but this is not any part of the role of "I". The corresponding reidentification *was* involved in the use of "*A*", and that makes an additional difference between them.

So perhaps "I" is not a name but rather another kind of expression indicating 'singular reference'. The logician's conception of the proper name after all only required *this* feature. There are expressions which logically and syntactically function as proper names without being names. Possibly definite descriptions do, and certainly some pronouns. "I" is called a pronoun, so we will consider this first. Unluckily the category 'pronoun' tells us nothing, since a singular pronoun may even be a variable (as in "If anyone says that, *he* is a fool") – and hence not any kind of singular designation of an object. The suggestion of the word "pronoun" itself is not generally borne out by pronouns. Namely, that you get the same sense in a sentence if you replace the pronoun in it by a name, common or proper: what name in particular, it would be difficult to give a general rule for. Perhaps "pronoun" seemed an apt name just for the personal pronouns and especially for "I". But the sense of the lie "I am not E.A." is hardly retained in "E.A. is not E.A.". So that suggestion is of little value.

Those singular pronouns called demonstratives ("this" and "that") are a clear example of non-names which function logically as names. For in true propositions containing them they provide reference to a distinctly identifiable subject-term (an object) of which something is predicated. Perhaps, then, "I" is a kind of demonstrative.

Assimilation to a demonstrative will not – as would at one time have been thought – do away with the demand for a conception of the object indicated. For, even though someone may say just "this" or "that", we need to know the answer to the question "this *what?*" if we are to understand him; and he needs to know the answer if he is to be meaning anything.[3]

Thus a singular demonstrative, used correctly, does provide us with a proper logical subject so long as it does not lack a 'bearer' or 'referent', and

[3] This point was not grasped in the days when people believed in pure ostensive definition without the ground's being prepared for it. Thus also in those days it was possible not to be so much impressed as we ought to be, by the fact that we can find no well-accounted-for term corresponding to "I" as "city" does to "London". It was possible to see that there was no 'sense' (in Frege's sense) for "I" as a proper name, but still to think that for each one of us "I" was the proper name of an 'object of acquaintance', a *this*. What *this* was could then be called "a self", and the word "self" would be felt to need no further justification. Thus, for example, McTaggart. See *The Nature of Existence* (Cambridge, 1921–7), Vol. II, ¶¶ 382, 386–7, 390–1, 394.

so it conforms to the logician's requirement for a name. And the answer to the question "this what?" might be taken to be "this self", if it can be shown that there are selves and that they are apparently what is spoken of by all these people saying "I". Thus would these philosophical inquiries about selves have a certain excuse.

It used to be thought that a singular demonstrative, "this" or "that", if used correctly, could not lack a referent. But this is not so, as comes out if we consider the requirement for an answer to "this what?" Someone comes with a box and says "This is all that is left of poor Jones." The answer to "this what?" is "this parcel of ashes"; but unknown to the speaker the box is empty. What "this" has to have, if used correctly, is something that it *latches on to* (as I will put it): in this example it is the box. In another example it might be an optical presentation. Thus I may ask "What's that figure standing in front of the rock, a man or a post?" and there may be no such object at all; but there is an appearance, a stain perhaps, or other marking of the rock face, which my "that" latches on to. The referent and what "this" latches on to may coincide, as when I say "this buzzing in my ears is dreadful", or, after listening to a speech, "That was splendid!" But they do not have to coincide, and the referent is the object of which the predicate is predicated where "this" or "that" is a subject.

There is no other pronoun but a demonstrative to which "I" could plausibly be assimilated as a singular term that provides a reference. Of course someone may say: "Why assimilate it at all? Each thing is what it is and not another thing! So 'I' is a pronoun all right, but it is merely the pronoun that it is." But that is no good, because 'pronoun' is just a rag-bag category; one might as well say: "It is the word that it is." The problem is to describe its meaning. And, if its meaning involves the idea of reference, to see what 'reference' is here, and how accomplished. We are now supposing that it is not accomplished as it is for a regular proper name; then, if "I" is not an abbreviation of a definite description, it must catch hold of its object in some other way – and what way is there but the demonstrative?

But there is a contrast between "I" and the ordinary demonstrative. We saw that there may be reference-failure for "this", in that one may mean "this parcel of ashes" when there are no ashes. But "I" – if it makes a reference, if, that is, its mode of meaning is that it is supposed to make a reference – is secure against reference-failure. Just thinking "I . . ." guarantees not only the existence but the presence of its referent. It guarantees the existence *because* it guarantees the presence, which is presence to consciousness. But note that here "presence to consciousness" means physical or real presence, not just that one is thinking of the thing. For if the thinking did not guarantee the presence, the existence of the referent could be doubted. For the same reason, if "I" is a name it cannot be an empty name. I's existence is existence in the thinking of the thought expressed by "I . . ." This of course is the point of the *cogito* – and, I will show, of the corollary argument too.

Whether "I" is a name or a demonstrative, there is the same need of a 'conception' through which it attaches to its object. Now what conception can be suggested, other than that of *thinking*, the thinking of the I-thought, which secures this guarantee against reference-failure? It may be very well to describe what selves are; but if I do not know that I am a self, then I cannot mean a self by "I".

To point this up, let me imagine a logician, for whom the syntactical character of "I" as a proper name is quite sufficient to guarantee it as such, and for whom the truth of propositions with it as subject is therefore enough to guarantee the existence of the object it names. He, of course, grants all that I have pointed out about the indirect reflexive. It cannot perturb him, so long as the 'way of being given' is of no concern to him. To him it is clear that "I", in my mouth, is just another name for E.A., "I" may have some curious characteristics; but they don't interest him. The reason is that "I" is a name governed by the following rule: 'If *X* makes assertions with "I" as subject, then those assertions will be true if and only if the predicates used thus assertively are true of *X*.'[4] This will be why Kripke – and others discussing Descartes – make the transition from Descartes' "I" to "*Descartes*".

Now first, this offers too swift a refutation of Descartes. In order to infer straight away that Descartes was wrong, we only need the information that Descartes asserted "I am not a body", together with the knowledge that he was a man: that is, an animal of a certain species; that is, a body living with a certain sort of life.

But there would and should come from Descartes' lips or pen a denial that, strictly speaking, *the man Descartes* made the assertion. The rule was sound enough. But the asserting subject must be the thinking subject. If you are a speaker who says "I", you do not find out what is saying "I". You do not for example look to see what apparatus the noise comes out of and assume that that is the sayer; or frame the hypothesis of something connected with it that is the sayer. If that were in question, you could doubt whether anything *was* saying "I". As, indeed, you can doubt whether anything is saying it out loud. (And sometimes *that* doubt is correct.)

Thus we need to press our logician about the 'guaranteed reference' of "I". In granting this, there are three degrees of it that he may assert.

(1) He may say that of course the user of "I" must exist, otherwise he would not be using "I". As he *is* the referent, that is what 'guaranteed reference' amounts to. In respect of such guaranteed reference, he may add, there will be no difference between "I" and "*A*". But the question is, why "I" was said to *refer* to the "I"-user? Our logician held that "I" was logically a proper name – a singular term whose role is to make a reference – for two reasons: one, that "I" has the same syntactical place as such expressions, and the other, that it can be replaced *salva veritate* by a (more ordinary) name of *X*

[4] My colleague Dr J. Altham has pointed out to me a difficulty about this rule about "I". How is one to extract the *predicate* for purposes of this rule in "I think John loves me"? The rule needs supplementation: where "I" or "me" occurs within an oblique context, the predicate is to be specified by replacing "I" or "me" by the indirect reflexive pronoun.

when it occurs in subject position in assertions made by *X*. In saying this, he no doubt thought himself committed to no views on the sense of "I" or what the "I"-user means by "I". But his second reason amounts to this: one who hears or reads a statement with "I" as subject needs to know whose statement it is if he wants to know of whom the predicate holds if the statement is true. Now, this requirement could be signalled by flashing a green light, say, in connection with the predicate, or perhaps adding a terminal '-O' to it. (I apologize to anyone who finds this suggestion altogether too fanciful, and beg him to suspend disbelief.) What would make such a signal or suffix into a referring expression? The essential argument cannot be an argument back from syntax to reference, for such an argument would depend only on the form of sentence and would be absurd. (e.g. no one thinks that "it is raining" contains a referring expression, "it".) And so it seems that our logician cannot disclaim concern with the sense of "I", or at any rate with what the "I"-user must mean.

(2) So the "I"-user must intend to refer to something, if "I" is a referring expression. And now there are two different things for "guaranteed reference" to mean here. It may mean (2a) guaranteed existence of the object meant by the user. That is to say, that object must exist, which he is taking something to be when he uses the expression in connection with it. Thus, if I suppose I know someone called "*X*" and I call something "*X*" with the intention of referring to that person, a guarantee of reference in this sense would be a guarantee that there is such a thing as *X*. The name "*A*" which I invented would have this sort of guaranteed reference. The "*A*"-user means to speak of a certain human being, one who falls under his observation in a rather special way. That person is himself, and so, given that he has grasped the use of "*A*", he cannot but be speaking of a real person.

If our logician takes this as an adequate account of the guaranteed reference of "I", then he will have to grant that there is a third sort of 'guaranteed reference', which "I" does *not* have. Guaranteed reference for that name "*X*" in this further sense (2b) would entail a guarantee, not just that there is such a thing as *X*, but also that what I take to be *X is X*. We saw that the "*A*"-user would not be immune to mistaken identification of someone else as '*A*'. Will it also be so with "I"?

The suggestion seems absurd. It seems clear that if "I" is a 'referring expression' at all, it has both kinds of guaranteed reference. The object an "I"-user means by it must exist so long as he is using "I", nor can he take the wrong object to be the object he means by "I". (The bishop may take the lady's knee for his, but could he take the lady herself to be himself?)

Let us waive the question about the sense of "I" and ask *only* how reference to the right object could be guaranteed. (This is appropriate, because people surely have here the idea of a sort of pure direct reference in which one simply first means and then refers to an object before one.) It seems, then, that this reference could only be sure-fire if the referent of "I" were both freshly defined with each use of "I", and also remained in view so long as something was being taken to be *I*. Even so there is an assumption that something else does not surreptitiously take its place. Perhaps we should say: such an assumption is extremely safe for "I", and it would be altogether an excess of scepticism to doubt it! So we accept the assumption, and it seems to follow that what "I" stands for must be a Cartesian Ego.

For, let us suppose that it is some other object. A plausible one would be *this body*. And now imagine that I get into a state of 'sensory deprivation'. Sight is cut off, and I am locally anaesthetized everywhere, perhaps floated in a tank of tepid water; I am unable to speak, or to touch any part of my body with any other. Now I tell myself "I won't let this happen again!" If the object meant by "I" is this body, this human being, then in these circumstances it won't be present to my senses; and how else can it be 'present to' me? But have I lost what I mean by "I"? Is that not present to me? Am I reduced to, as it were, 'referring in absence'? I have not lost my 'self-consciousness'; nor can what I mean by "I" be an object no longer present to me. This both seems right in itself, and will be required by the 'guaranteed reference' that we are considering.

Like considerations will operate for other suggestions. Nothing but a Cartesian Ego will serve. Or, rather, a stretch of one. People have sometimes queried how Descartes could conclude to his *RES cogitans*.[5] But this is to forget that Descartes declares its essence to be nothing but thinking. The thinking that thinks this thought – that is what is guaranteed by "cogito".

Thus we discover that *if* "I" is a referring expression, then Descartes was right about what the referent was. His position has, however, the intolerable difficulty of requiring an identification of the same referent in different "I"-thoughts. (This led Russell at one point to speak of 'short-term selves'.)

Our questions were a combined *reductio ad absurdum* of the idea of "I" as a word whose role is to 'make a singular reference'. I mean the questions how one is guaranteed to get the object right, whether one may safely assume no unnoticed substitution, whether one could refer to oneself 'in absence', and so on. The suggestion of getting the object right collapses into absurdity when we work it out and try to describe how getting hold of the wrong object may be excluded.

How, even, could one justify the assumption, if it is an assumption, that there is just one thinking which is this thinking of this thought that I am thinking, just one thinker? How do I know that 'I' is not ten thinkers thinking in unison? Or perhaps not quite succeeding. That might account for the confusion of thought which I sometimes feel. – Consider the reply "Legion, for we are many", given by the possessed man in the gospel. Perhaps we should take that solemnly, not as a grammatical joke.[6] These

[5] For example A. J. Ayer. See *Language, Truth and Logic* (2nd edn, London, 1946), p. 142.

[6] Ambrose Bierce has a pleasant entry under "I" in the *Devil's Dictionary*: "I is the first letter of the alphabet, the first word of the language, the first thought of the mind, the first object of the affections. In grammar it is a pronoun of the first person and singular number. Its plural is said to be *We*, but how there can be more than one myself is doubtless clearer to the grammarians

considerations refute the 'definite description' account of "I". For the only serious candidate for such an account is "The sayer of this", where "sayer" implies "thinker".

Getting hold of the wrong object *is* excluded, and that makes us think that getting hold of the right object is guaranteed. But the reason is that there is no getting hold of an object at all. With names, or denoting expressions (in Russell's sense) there are two things to grasp: the kind of use, and what to apply them to from time to time. With "I" there is only the use.

If this is too hard to believe, if "I" *is* a 'referring expression', then Descartes was right. But now the troubles start. At first, it seems as if what "I" stands for ought to be the clearest and certainest thing – what anyone thinking of his own thinking and his own awareness of anything is most evidently aware of. It is most certain because, as Augustine said, it is involved in the knowledge of all mental acts or states by the one who has them. They could not be doubted. But the *I*, the 'mind', the 'self', was their subject, not their object, and looking for it as an object resulted, some people thought, in total failure. It was not to be found. It was rather as it were an area of darkness out of which light shone on everything else. So some racked their brains over what this invisible subject and the 'thinking of *it*' could be; others thought there was no such thing, there were just all the objects, and hence that "I", rather, was the name of the whole collection of perceptions. But that hardly fitted its grammar, and anyway – a problem which utterly stumped Hume – by what was *I* made into a unity? Others in effect treat selves as postulated objects for "I" to be names of in different people's mouths. Yet others denied that the self was invisible, and claimed that there is a unique feeling of oneself which is indescribable but very, very important, especially in psychology, in clinical psychology, and psychiatry.

With that thought: "The *I* was subject, not object, and hence invisible", we have an example of language itself being as it were possessed of an imagination, forcing its image upon us.

The dispute is self-perpetuating, endless, irresoluble, so long as we adhere to the initial assumption, made so far by all the parties to it: that "I" is a referring expression. So long as that is the assumption you will get the deep division between those whose considerations show that they have not perceived the difficulty – for them "I" is in principle no different from my "*A*"; and those who do – or would – perceive the difference and are led to rave in consequence.

And this is the solution: "I" is neither a name nor another kind of expression whose logical role is to make a reference, *at all*.

Of course we must accept the rule "If *X* asserts something with 'I' as subject, his assertion will be true if and only if what he asserts is true of *X*." But if someone thinks that is a sufficient account of "I", we must say "No, it

than it is to the author of this incomparable dictionary. Conception of two myselves is difficult, but fine. The frank yet graceful use of "I" distinguishes a good author from a bad; the latter carries it with the manner of a thief trying to cloak his loot."

is not", for it does not make any difference between "I" and "*A*". The truth-condition of the whole sentence does not determine the meaning of the items within the sentence. Thus the rule does not justify the idea that "I", coming out of *X*'s mouth, is another name for *X*. Or for anything else, such as an asserting subject who is speaking through *X*.

But the rule does mean that the question "*Whose* assertion?" is all-important. And, for example, an interpreter might repeat the "I" of his principal in his translations. Herein resides the conceivability of the following: someone stands before me and says, "Try to believe this: when I say "I", that does not mean this human being who is making the noise. I am someone else who has borrowed this human being to speak through him." When I say "conceivability" I don't mean that such a communication might be the truth, but only that our imagination makes something of the idea. (Mediums, possession.)

If I am right in my general thesis, there is an important consequence – namely, that "I am E.A." is after all not an identity proposition. It is connected with an identity proposition, namely, "This thing here is E.A.". But there is also the proposition "I am this thing here".

When a man does not know his identity, has, as we say, 'lost his memory', what he doesn't know is usually that *that* person he'd point to in pointing to himself (this is the direct reflexive) is, say, Smith, a man of such-and-such a background. He has neither lost the use of "I", nor would he feel at a loss what to point to as his body, or as the person he is; nor would he point to an unexpected body, to a stone, a horse, or another man, say. The last two of these three points may seem to be part of the first of them; but, as we have seen, it is possible at least for the imagination to make a division. Note that when I use the word "person" here, I use it in the sense in which it occurs in "offences against the person". At this point people will betray how deeply they are infected by dualism, they will say: "You are using 'person' in the sense of 'body'" – and what *they* mean by "body" is something that is still there when someone is dead. But that is to misunderstand "offences against the person". None such can be committed against a corpse. 'The person' is a living human body.

There is a real question: with what object is my consciousness of action, posture and movement, and are my intentions connected in such fashion that *that* object must be standing up if I have the thought that I am standing up and my thought is true? And there is an answer to that: it is this object here.

"I am this thing here" is, then, a real proposition, but not a proposition of identity. It means: this thing here is the thing, the person (in the 'offences against the person' sense) of whose action *this* idea of action is an idea, of whose movements *these* ideas of movement are ideas, of whose posture *this* idea of posture is the idea. And also, of which *these* intended actions, if carried out, will be the actions.

I have from time to time such thoughts as "I am sitting", "I am writing",

"I am going to stay still", "I twitched". There is the question: in happenings, events, etc. concerning what object are these verified or falsified? The answer is ordinarily easy to give because I can observe, and can point to, my body; I can also feel one part of it with another. "This body is my body" then means "My idea that I am standing up is verified by this body, if it is standing up". And so on. But observation does not show me which body is the one. Nothing shows me that.[7]

If I were in that condition of 'sensory deprivation', I could not have the thought "this object", "this body" – there would be nothing for "this" to latch on to. But that is not to say I could not still have the ideas of actions, motion, etc. For these ideas are not extracts from sensory observation. If I do have them under sensory deprivation, I shall perhaps *believe* that there is such a body. But the possibility will perhaps strike me that there is none. That is, the possibility that there is then nothing that I am.

If "I" were a name, it would have to be a name for something with this sort of connection with this body, not an extra-ordinary name for this body. Not a name for this body because sensory deprivation and even loss of consciousness of posture, etc., is not loss of *I*. (That, at least, is how one would have to put it, treating "I" as a name.)

But "I" is not a name: these I-thoughts are examples of reflective consciousness of states, actions, motions, etc., not of an object I mean by "I", but of this body. These I-thoughts (allow me to pause and think some!) . . . are unmediated conceptions (knowledge or belief, true or false) of states, motions, etc., of this object here, about which I can find out (if I don't know it) that it is E.A. About which I did learn that it is a human being.

The I-thoughts *now* that have *this* connection with E.A. are I-thoughts on the part of the same human being as the I-thoughts that had that connection twenty years ago. No problem of the continuity or reidentification of 'the *I*' can arise. There is no such thing. There is E.A., who, like other humans, has such thoughts as these. And who probably learned to have them through learning to say what she had done, was doing, etc. – an amazing feat of imitation.

Discontinuity of 'self-feeling', dissociation from the self-feeling or self-image one had before, although one still has memories – such a thing is of course possible. And so perhaps is a loss of self-feeling altogether. What this 'self-feeling' is is no doubt of psychological interest. The more normal state is the absence of such discontinuity, dissociation and loss. That absence can therefore be called the possession of 'self-feeling': I record my suspicion that this is identifiable rather by consideration of the abnormal than the normal case.

Self-knowledge is knowledge of the object that one is, of the human animal that one is. 'Introspection' is but one contributory method. It is a rather doubtful one, as it may consist rather in the elaboration of a self-image than in noting facts about oneself.

[7] Professor Føllesdal and Mr Guttenplan tell me that there is some likeness between what I say and what Spinoza says. I am grateful for the observation; but cannot say I understand Spinoza.

If the principle of human rational life in E.A. is a soul (which perhaps can survive E.A., perhaps again animate E.A.) *that* is not the reference of "I". Nor is it what I am. I am E.A. and shall exist only as long as E.A. exists. But, to repeat, "I am E.A." is not an identity proposition.

It will have been noticeable that the I-thoughts I've been considering have been only those relating to actions, postures, movements and intentions. Not, for example, such thoughts as "I have a headache", "I am thinking about thinking", "I see a variety of colours", "I hope, fear, love, envy, desire", and so on. My way is the opposite of Descartes'. These are the very propositions he would have considered, and the others were a difficulty for him. But what were most difficult for him are most easy for me.

Let me repeat what I said before. I have thoughts like "I am standing", "I jumped". It is, I said, a significant question: "In happenings, events, etc., concerning what object are these verified or falsified?" – and the answer was: "this one". The reason why I take only thoughts of actions, postures, movements and intended actions is that only those thoughts both are unmediated, non-observational, and also are descriptions (e.g. "standing") which are directly verifiable or falsifiable about the person of E.A. Anyone, including myself, can look and see whether that person is standing.

That question "In happenings, events, etc., concerning what object are these verified or falsified?" could indeed be raised about the other, the Cartesianly preferred, thoughts. I should contend that the true answer would be "if in any happenings, events, etc., then in ones concerning this object" – namely the person of E.A. But the description of the happenings, etc., would not be just the same as the description of the thought. I mean the thought "I am standing" is verified by the fact that this person here is *standing*, falsified if she is not. This identity of description is entirely missing for, say, the thought "I see a variety of colours". Of course you may say, if you like, that this is verified if this person here sees a variety of colours, but the question is, what is it for it to be so verified? The Cartesianly preferred thoughts all have this same character, of being far removed in their descriptions from the descriptions of the proceedings, etc., of a person in which they might be verified. And also, there might not be any. And also, even when there are any, the thoughts are not thoughts of such proceedings, as the thought of standing is the thought of a posture. I cannot offer an investigation of these questions here. I only want to indicate why I go after the particular "I"-thoughts that I do, in explaining the meaning of "I am E.A." This may suffice to show why I think the Cartesianly-preferred thoughts are not the ones to investigate if one wants to understand "I" philosophically.

Suppose – as is possible – that there were no distinct first-person expression, no pronoun "I", not even any first-person inflection of verbs. Everyone uses his own name as we use "I". (Children sometimes do this.) Thus a man's own name takes the place of "I" in this supposed language. What then? Won't his own name still be a name? Surely it will! He will be using what is syntactically *and* semantically a name. That is, it is semantically

a name in other people's mouths. But it will not be so in his mouth, it will not signify like a name in his utterances.

If I used "E.A." like that, and had no first-person inflections of verbs and no such words as "I", I should be in a difficulty to frame the proposition corresponding to my present proposition: "I am E.A." The nearest I could get would be, for example, "E.A. is the object E.A." That is, "E.A. is the object referred to by people who identify something as E.A."

There is a mistake that it is very easy to make here. It is that of supposing that the difference of self-consciousness, the difference I have tried to bring before your minds as that between "I"-users and "*A*"-users, is a private experience. That there is this asymmetry about "I": for the hearer or reader it is in principle no different from "*A*"; for the speaker or thinker, the "I"-saying subject, it is different. Now this is not so: the difference between "I"-users and "*A*"-users would be perceptible to observers. To bring this out, consider the following story from William James. James, who insisted (rightly, if I am right) that consciousness is quite distinct from self-consciousness, reproduces an instructive letter from a friend: "We were driving . . . in a wagonette; the door flew open and *X*, alias 'Baldy', fell out on the road. We pulled up at once, and then he said 'Did anyone fall out?' or 'Who fell out?' – I don't exactly remember the words. When told that Baldy fell out he said 'Did Baldy fall out? Poor Baldy!'"[8]

If we met people who were *A*-users and had no other way of speaking of themselves, we would notice it quite quickly, just as his companions noticed what was wrong with Baldy. It was not that he used his own name. That came afterwards. What instigated someone to give information to him in the form "Baldy fell out" was, I suppose, that his behaviour already showed the lapse of self-consciousness, as James called it. He had just fallen out of the carriage, he was conscious, and he had the idea that someone had fallen out of the carriage – or he knew that someone had, but wondered who! That was the indication of how things were with him.

Even if they had spoken a language without the word "I", even if they had had one without any first-person inflexion,[9] but everybody used his own name in his expressions of self-consciousness, even so, Baldy's conduct would have had just the same significance. It wasn't that he used "Baldy" and not "I" in what he said. It was that his thought of the happening, falling out of the carriage, was one for which he looked for a subject, his grasp of it one which required a subject. And that could be explained even if we didn't have "I" or distinct first-person inflexions. He did not have what I call 'unmediated agent-or-patient conceptions of actions, happenings and states'. These conceptions are subjectless. That is, they do not involve the connection of what is understood by a predicate with a distinctly conceived subject. The (deeply rooted) grammatical illusion of a subject is what generates all the errors which we have been considering.

[8] *Principles of Psychology*, II (London 1901), p. 273 n.

[9] In Latin we have "ambulo" = "I walk". There is no subject-term. There is no need of one.

3 Substance

The raising of certain difficulties about the notion of substance belongs especially to the British Empiricist – that is to say our – tradition. We can see a starting-point for them in Descartes' considerations about the wax in the Second Meditation. Descartes concluded there that it was by an act of purely intellectual perception that we judge the existence of such a thing as *this* wax – a doctrine the meaning of which is obscure.

Let me sketch at least some of the troubles that have been felt on this subject. First, there is the idea of the individual object. What sort of idea is that, and how got? This individual object is the same – 'persists' as we say – through many changes in its sensible properties, or sensible appearances; what is the individual itself all this time? Second, supposing that question should be answered in the case in hand by 'It is wax", is it not one objection to this answer that it gives a *general* term "wax" as an answer to the question "What is the individual?" Surely we wanted to know: what is the individual thing *qua* individual, in its individuality? And this cannot be answered by giving a predicate which not merely logically can be true of many individuals, but does actually fail to mark out this one from others. Next, even accepting this answer: "It is wax", what can being wax be except: being white and solid at such and such temperatures, melting at such and such temperatures . . . etc., etc.? Are not the ideas of kinds of substances given by more or less arbitrary lists chosen from the properties found by experience to go together? In that case, the general idea 'wax' will be equivalent to the chosen list; and the particular, individual, parcel of wax is at any time the sum of its sensible appearances. Any other notion of substance surely commits us on the one hand to unknowable real essences, and on the other to an unintelligible 'bare particular' which underlies the appearances and is the subject of predication but just for that reason can't *in itself* be characterized by any predicates. This picture of the appearances or the properties as a sort of clothes reminds one of Butler's lines about Prime Matter:

> He had first matter seen undrest;
> He took her naked, all alone,
> Before one rag of form was on.

The picture of substance is too unacceptable, so following Russell we must speak of 'bundles of qualities' or following Ayer of 'totalities of appearances' which are unified not by their relation to some further entity but by their own interrelations. We would rather not admit anything so doubtful as that

From *Proceedings of the Aristotelian Society*, supplementary volume, 38 (1964).

'obscure and relative' idea of substance which Locke said the ideas of qualities, actions and powers brought with them: the idea, namely, of the substrate which supports them. (Indeed I believe that the sort of account you get in H. W. B. Joseph of the ultimate characterless subject of predication is a conflation of Locke on substance and (an attempt to understand) Aristotle on the matter of substantial change – the stuff which is not *as such X*, and not not-*X* either, whatever *X* may be.)

These were standard arguments and opinions in the times when I was growing up, and they are probably still extremely familiar. One doctrine closely associated with them, which has been criticized before but is in constant need of rebuttal, is the view that the *individual* has no 'nominal essence', that the proper name either lacks all connotation, or has that of a more or less arbitrarily incomplete description giving the history of the individual. This doctrine indeed concerns all proper names, not just those of substances, but it does concern substances too. About this I have written elsewhere, and here will merely repeat that to describe a word as a proper name at all is a great deal of information about its sense, which only needs completion by saying of what kind of thing it is a proper name. The doctrine that individuals have nothing that is essential to them suggests a phantasmic notion of the individual as a 'bare particular' with no properties, because it supposes a continued identity independent of what is true of the object. This was thought to be *the* notion of substance, to which the objections were well known.[1] One of the considerations brought forward in erecting this notion (for it is not a straw man, real humans *have* gone in for it) seems so idiotic as to be almost incredible, namely that the substance is the entity that has the properties, and so it itself has not properties. Philosophers have been divided between those who defend some such notion as necessary and those who reject substance because it involves this notion and is therefore itself absurd.

The lump of wax melts and becomes liquid. The argument – if it can be called that – for the propertyless subject would suggest that the subject of the properties 'melting' and 'liquidity' is the individual which in itself has no properties. This must depend on taking "in itself" to mean "apart from having properties". But another possible meaning of "in itself" does not lead to the characterless substrate which people supposed was meant by "substance". "What a thing has in itself" might mean "What is always and necessarily true of it". Descartes' argument was that the wax must be something grasped by the intelligence because all the sensible properties changed but it was the same wax. Now this argument does not require a propertyless subject, but a subject with some permanent properties – which, however, he says are not 'sensible properties'.

What does Descartes mean by sensible properties? He mentions colour,

[1] Alas, this belief is not so much a thing of the past as I supposed. See for example A. H. Basson's *David Hume* (London, 1958), pp. 136 ff., and J. P. Griffin's *Wittgenstein's Logical Atomism* (Oxford, 1964), p. 71 – the ink of this one is scarcely dry.

shape, size, being liquid, being hot, yielding no sound when rapped. After heating, "Whatever fell under taste, smell, sight, touch, or hearing has now changed".

There are relevant differences between the properties mentioned. Visible colour, shape and size are not substance-involving. 'The sound it gives out if you rap it' seems to be substance-involving in this sense: the question arises "If you rap what?" and the answer is "the wax". But sound itself of course is not substance-involving. What I mean by "not substance-involving" is this: you can suppose a man to see a coloured expanse without there having to be any substance (or, of course, collection of substances) *whose* expanse, or part of whose expanse, it is. One of the problems of epistemology that first strikes one – did first strike me – is: how do I know the things I look at have behinds? Why shouldn't they have the sort of merely phenomenal existence a rainbow has? This question arises because colour, together with its determinations of shape and size, is not substance-involving. Just this is what I take a philosopher to be driving at if he says things like "All I have got when I look, and, as I say, see a red curtain, is a visual content specifiable as light and dark and colour patches thus and so arranged." I do not think he is effectually answered by deriding him for saying by implication that he does not, strictly speaking, see a red curtain hanging in folds.

The properties known as secondary qualities in modern philsophy have a claim to be called "sensible" in a much more restricted sense than that in which we could say malleability was sensible. To receive impressions of secondary qualities, you merely have to let the appropriate sense-organ be affected; that is why one can always imagine that the quality is a *mere* sense content. This is indeed a lot easier to imagine for 'white' than for 'soft' but it can be imagined for all those qualities.

But no list only of *those* sensible properties would be adequate to comprise the idea – the 'nominal definition' – of a particular substantial kind. There are always further properties, such as malleability and melting at 44 °C, which, though eminently ascertainable by the senses, are substance-involving properties. The red patch you see might have – could be imagined to have – only the sort of existence that a rainbow does. If I asked someone to see if the rainbow was malleable or melted at 44 °C, this would imply a conception of a rainbow as made of a kind of stuff.

Let us now consider the reasonableness of defining a substance as the totality of its appearances. "The appearances" of a substance suggests its sensible properties in the restricted sense, i.e. the secondary qualities, together with their qualifications of size, shape and mutual arrangement.

The reason is this: usually our judgements of what is there are right, and then we don't have to bother about appearances; but when they are wrong we can retreat to appearances. This retreat may consist merely in saying "It *looked* as if there were a fly on the painting, but actually there was no fly" (where the appearance is one of a fly), or, extending the notion of 'appearance' to the sense of touch, "It *felt* as if there were fur there in the hole

where I put my hand – but there was not". Here the appearance is one of fur: an appearance to the sense of touch, as that of the fly was to the sense of sight.

Now when there are such appearances, the secondary qualities involved are usually not mere *appearances* of secondary qualities. In a *trompe l'oeil* painting the colours aren't a mere appearance; rather they, in their disposition and relation to the surroundings, are what yields, or at least are a necessary element in, the appearance of there being a fly on the painting of a lily, or an archway in a wall leading into another room. And the felt softness bears the same sort of relation to the incorrect judgement "There's fur there" as the colour patches do to the incorrect judgement "There is a doorway there". If one did not know what was there, but knew only – no matter how – that the judgement one was inclined to make, "That's a fly" (or, "It's fur") was wrong, one could retreat to the description of the colour patches (the texture) as what one saw (felt) *in that* one was inclined to think one saw a fly (felt fur). And there does not in fact have to be illusion or incorrect judgement to *entitle* one to make this sort of retreat; that type of case merely *forces* one to retreat and that is the use of considering it.

For these reasons, quite generally there is I think no objection to calling the 'secondary qualities' (with their immediate qualifications) "appearances" of the things we incline to think are there when perceiving the qualities. But malleability, though a sensible property, is not like this: it is not in any case an *appearance* of the malleable thing. There can of course be an appearance of malleability in the sense that someone could make it look as if something were malleable when it was not; but that does not mean that "malleability" is itself a word for an appearance – for a way things strike the senses.

We can see three ranks of predicate that apply to substances; the substantial ones themselves, like "alive", "horse", "gold"; the predicates that are not substantial but are substance-involving like "malleable", "in powder form", "awake"; and predicates that are neither substantial nor substance-involving. These are the secondary-quality words, together with such qualifications as go with them.

To repeat, if I asked you to see if the rainbow would melt at 44 °C, this would imply a conception of a rainbow as composed of stuff, so that a sample of it could be brought away and subjected to tests. "Malleability" means that the stuff can be beaten into a shape which it will then retain if not further interfered with. Thus you could not consider whether something was malleable unless you had the conception of a lump of stuff whose properties could be further investigated – but that is already to have a partial conception of substance. Thus, though malleability is obviously a sensible property, nevertheless a thoroughgoing phenomenalist would want to analyse it out, just as he would want to analyse the substantial predicates out.

Substantial predicates are more than substance-involving. They tell you what kind or kinds of substance *that lump of stuff* is. Something must be *that lump of stuff* in order to be so much as a candidate for having malleability.

This makes it sound as if it were normal to pass from the bare characterization "lump of stuff" straight to the enquiry into substance-involving predicates. While this may happen, it is not normal; the lump is generally already known to be a lump of stuff of a certain kind – a bit of copper, say – and the kind is told you more or less specifically by the substantial predicates. Very many substantial predicates enter naturally into the characterization of appearances – it felt like fur, it looked like metal. It is notorious that these characterizations are often irreducible. If it felt like fur, it felt soft; but with the quite peculiar softness characteristic of this or that kind of fur.

But the fact that something looks, smells, tastes, feels and sounds like *X* – or as many of these as is possible – does not prove that it is *X*: for all this is appearance, capable of conflicting with reality. For example it may not have the right origin or chemical structure or reactive properties to be *X*. Predicates expressing a substance's origin are likely to be and predicates expressing its chemical structure are certainly themselves substantial predicates.

Some people will want to know why the secondary qualities are not substance-involving. I said that the red that you see might have an existence like that of the rainbow; it might be a colour to be seen from this position, but not the colour of any substance. It may be said to me: "Suppose you *know quite well* that it is the colour of the red plate you are looking at. You see what evidently is a plain red plate of uniform colour. The supposition that this is simply a red patch, to be seen when you look in that direction, with no more substantial existence than a rainbow has, is plain ridiculous: you know there is no question of it. The red is the red of the *plate*, just as that other bit of red is the red of the *curtain*. The plate and the curtain are of course not hidden, not occult substrates which you cannot know; but they imply a substantial existence, and reference to them enters into your account of the expanses of red that you see. So aren't colours also substance-involving, when they are the colours of objects? Admittedly this is not the same manner of substance-involvement as malleability has; but, from the fact that not all the red that you may see has to be the red *of* anything, you have inferred that no red that you see can be immediately perceived to be the red of something; and that is wildly unjustified."

I fear that this involves us in a fresh start. Descartes defined substance as what needs nothing else (apart from divine co-operation) in order to exist; but if he were faced with the objection that many substances need oxygen, or a certain temperature, to exist, he would no doubt say he had not meant that; so presumably he meant something like Aristotle, who defined individual substance as what exists without either being predicated of or existing in anything else.

Now let us consider the red patches of Cambridge twentieth-century philosophy ("I see a red patch" seemed to be very clear, very certain, very safe), and ask: Are these substances by the Aristotelico–Cartesian account?

They were certainly supposed to be individuals, particulars – should we then say: they ought to have been conceived as substances, if a substance has independent existence, i.e. exists not in another thing?

The answer to this will depend on (1) whether these red patches, assumed to be real entities, are supposed to be pure objects of sense and (2) whether they are thought not merely to have an *esse* which is *percipi* but also to exist essentially in dependence on an act of *percipere* by a mental substance. I am only concerned with the former question, which I think will be explained, if not resolved, by the following:

If what I am looking at is a plain red plate, then there is before me and in plain view an expanse of a standing red colour. But I see a certain variation produced by the shadowing of some part of the plate, which is not flat but has a slope up to its edge. If I look carefully I will see a lot of variation in the appearance of the surface, very light spots and tiny short streaks, some a matter of particles of dust settled there, some a matter of minute variations of light and shadow. Also I see high-lights. Yet I say with confidence that this is a uniformly red plate. I learn to say "red plate" "white door" without regard to shines and streaks and variations of light and dark. Now if I speak of the red patch that I see, is *the part where the high-light is* part of the red patch? That part of the expanse is of a standing red colour, and I say I see something of a uniform red colour. But if where I see not red but a high-light is part of the red patch that I see, then that red patch is not something whose *esse* is *percipi*. Its red colour is a standing red colour, not seen by me in all parts, although I see the whole expanse in question. Then it is not a substance by the Aristotelico–Cartesian definition: its identity is that of the standing colour *of* that part of the surface of a plate, and its existence is *in* something else.

Now for a phenomenalist *this* red patch, which = the expanse of plate visible to me, is just as much a construct, an inference, as the plate itself. The fact that the high-light moves about on the plate when one moves one's head or moves the plate proves that it is just a *shine*; but that it *will* so move can't be *seen* in seeing it, but only judged or guessed on some grounds – and the grounds must be the way it now looks together with my past experience.

Is there not a description which gives simply what is *seen* – and does not depend on whether one or another thing which can't possibly be being seen is the case? Certainly one sees a plate; yet it isn't a plate if it has no behind, and one doesn't see in one viewing that it has a behind. So surely "a plate" is simply a true description of what what one sees in fact *is*, and also perhaps one's straight-off description of what it strikes one as; but yet in *a* sense of "sees" there is more than can be *seen* to calling it a plate – as in the case of John Austin's "To-day I saw a man born in Jerusalem" (said in Oxford). Of course there isn't a born-in-Jerusalem look about a man, so that case is particularly clear – but aren't the two cases essentially similar? We are merely distracted by the fact that there is no born-in-Jerusalem look about a man – whereas there is a plate-look about. . . .

About *what?* – Here one wants to say something like: the *red patch as seen* – the one that isn't red where the high-lights are and that is variegated by shadows and all sorts of spots and streaks.

But even this red patch is not one whose *esse* is *percipi*, unless one can be supposed not to notice the true character of that whose *esse* is *percipi*, to observe it more closely, to realize one had mistaken it. For the realization of the spots and streaks and shadows and high-lights is a gradual process of discovery.

Locke:

> When we set before our eyes a round globe of any uniform colour, e.g. gold, alabaster or jet, it is certain that the idea thereby imprinted in our mind is of a flat circle variously shadowed, with several degrees of light and brightness coming to our eyes. But we having by use been accustomed to perceive what kind of appearance convex bodies are wont to make in us, what alterations are made in the reflections of light by the difference of the sensible figures of bodies, the judgment presently, by an habitual custom, alters the appearances into their causes: so that, from that which truly is variety of shadow or colour collecting the figure, it makes it pass for a mark of figure, and frames to itself the perception of a convex figure and an uniform colour; when the idea we receive from thence is only a plane variously coloured, as is evident in painting.

The notion of the 'idea' as Locke calls it (or visual sensation, impression, experience or datum, as later writers have called the same thing) is in this context, I suggest, a conflation of disparate notions: *What I have the impression of seeing* – which may be quite properly said to be "a globe" or "a red plate" – and something quite different and very difficult to get at, which we want to call the "*purely visual*" about what is seen. It is what you'd get if, adopting the suggestion of Leonardo, you held up a glass pane vertically before you when you were looking straight ahead and supposed to be painted on it with utter accuracy *exactly* the colour behind it, as seen, in every part of it. The result represents what is thought of as the minimal, uninterpreted visual impression, which is the basis of all else. And it seems as though in this conception the difference between the objective and the subjective appearance – between the highlight, or colour changed by the light it is seen in, on the one hand, and the drug-induced or astigmatic colours and perspectives, on the other – is quite unimportant. But this pane would in turn be only an ordinary object of perception: it does duty for something else; it merely carries what *has to be understood as a picture of* a purely visual object.

4 The Subjectivity of Sensation

Doctors match blood samples with colours on a scale in order to judge cursorily how far someone is anaemic. This fact, and similar ones, are of great significance for anyone who wants to understand the metaphysical status of colours and other 'secondary qualities'.

Someone might complain that the procedure is rather subjective. He would presumably mean: the thing is a mere matter of an impression, and so uncheckable and unreliable. One man may have one impression, another another, without any way of ensuring correctness; without, indeed, there being any such thing as correctness. For colour is essentially an object of sight, a quality of a sensory impression, and where there is no eye to see there is no colour, except in the sense of power in an object to produce an impression.

"The procedure (of the doctor) is essentially subjective, in the sense that the impression depends on the disposition of the subject, and, just *qua* impression, any one has the same status as any other."

And yet there is such a thing as competence here, and the competent won't disagree. One man might have reason to mistrust his own judgement, and might ask another to check it, without any feeling of absurdity.

Herein resides our problem. For one man cannot ask another to check his judgement, except on the supposition that the matter of judgement is accessible to the other. Neither can he check his own judgement, except on the assumption that he can tell what he is judging about is the same. We see here the reason for the so-called 'incorrigibility' of descriptions of impressions, except under the claim that one had not meant what one said. For, suppose I call something blue, a china circle on a clock face, let us say. Someone protests that it is not blue. "Well," I say, "it is perhaps a greenish blue, but to me that seems blue." He retorts that on the contrary it is yellow. Surprised at this, I go up to the clock and look at it again and now I agree I was mistaken: the china circle is yellow. But can I say I was wrong in giving the colour 'blue' as the colour it had struck me as having? I'm now attending to the impression I have of it now and that is not the same impression. Could I say: "It strikes me now as striking me as having the colour that it struck me as having before, and it now strikes me as yellow, therefore I was wrong in thinking before that it struck me as blue"? This would be unintelligible, if at least I grant that my former thought "It's blue" was a perceptual judgement. For that is to say it involved an expression of colour-experience on my part, which, as I meant it, gave the colour that the object struck me as having. If I meant blue when I said "blue", then could I not also

From *Ajatus* (Yearbook of the Philosophical Society of Finland), 36 (1976).

have given the colour I was calling "blue" as a paradigm of blue – of what I meant by "blue" at least to myself?

If one takes another look and gives a different, incompatible, description of colour, then *either* what one is describing is the same, but is not an impression; *or*, if it is an impression, it is a new impression, and so one is using a new impression to correct the description of the thing seen, not to correct the description of former impressions. It might be objected to this that one may not be just responding to a current impression, but also making a memory-judgement that the present impression is an impression of the same colour (say), as an earlier impression was of. One knows that one took the earlier impression to be an impression of blue, and so one judges that one was wrong. But could one recall the experience itself – recall the previous impression actually seeming to be an impression of blue? If so, how could the present one seem not to be of blue and yet seem to be of the same colour? At most, then, one could say "I can't recapture the earlier impression, I know I said 'blue' and I have no impression of a difference, and now I don't want to say 'blue'". For if one claims to recapture the earlier impression, must one not think that one had previously misnamed the colour, not really meant "blue"? One will then be correcting, not one's former idea of the colour the thing seemed to be, but rather one's name for it.

Contrast mistaken identification of a person. This has two forms, illustrated by the following story: I knew of two people, *P* and *M*, whom I had not yet encountered. When I met them something caused me to associate the name "*P*" with the man *M*; thereafter for a while I recognized *M* when I saw him, but thought of him as *P*. I meant the wrong man by the name, though so far as I know I reidentified the man quite correctly: that is, I correctly recognized the man that I took to be *P*. Later still I learned my mistake and got the name "*P*" attached to the right man. But one day someone quite different called "*S*" came to see me; he looked rather like *P* and I took him for *P* and so introduced him to someone who was there. Thus the two forms of misidentification are these: with one there is a wrong association of name and person; with the other, there is the right association of name and person, but one wrongly takes *this* man, who stands before one, to be that person.

The mistaken application of the name of a colour – when we mean by this a colour as experienced, the colour one is having an impression of – has only the first form.

Someone may say: could one not know in a general sort of way what colour magenta is, for example, and now take *this* colour, wrongly, to be magenta? It is true: that is possible. But 'know in a general way' here means 'have a rough idea'. Magenta is a brilliant colour narrowly located somewhere in the range of purples, but I am vague about its exact place within that range. I need a paradigm to confirm or correct myself, i.e. to supply the meaning of the word. That means that when I 'think the impression is of magenta', what I think is that if I had hold of a paradigm of

magenta so as to be sure what colour magenta was, I would still have said: the impression is of magenta. So this example does not supply a counter-example.

I am uncertain how far this 'incorrigibility of descriptions of impressions' extends. As I have insisted elsewhere, "impression" or "sense-impression" is a term of very wide application. One may for example have a visual impression of a certain sort of continuity in a pattern, of numerical equality in sets of elements, of a face distorted with rage, of motion in an unmoving picture; of causality; one may have a complex impression – composed of elements available to more than one sense – of someone's being a bit put out (*verstimmt*). And one may be discontented with one's verbal expressions of one's impressions and so want to correct them. Furthermore, inconsistency in descriptions of an impression which may be discovered may or may not strike one as a ground for rejecting one of them. I am concerned with none of these interesting questions: I am concerned only with impressions such as may be described in pure appearance-terms, by which I mean terms for 'proper sensibles', as Aristotle called them, or what modern philosophy has called secondary qualities.

For the doctors, with whom I started, make judgements of identity and difference of colours. So they are judging in the domain of proper sensibles or secondary qualities. Their judgements are the expression of sensory impressions; what they are judging is an object, colour, which is the content of a sensory impression.

And yet the doctor asks his assistant, say, to check whether the colours match as he thinks they do. So he is treating the matter of judgement as not changing from one inspection to another.

This offers us a suggestion about what it means for an impression – an impression of a proper sensible – to be right. It suggests the possibility that, contrary to Cartesian assumptions, that 'incorrigibility' which we have been discussing is not any kind of correctness. For here at any rate an impression's being correct is its giving correct information about something else: in this case what I will call "standing colours of objects". To repeat, in this case the judgement is not a judgement about something non-sensory *via* something sensory, it is indeed about the proper object of sight. But what some people have believed about these objects, that their *esse* is *percipi*, i.e. that the proper object of sight does not exist except as seen and when it is being seen, is contradicted by the validity of the doctor's procedure.

One finds, or may construct, arrangements of colour which bring it out that the colour appearing to one who sees depends on the background colours: that it varies with variation of background. An example appears on the front cover of Robert Schwartz's collection on perception and knowledge: it shows two red squares traversed by stripes, which are green in one case while in the other they are blue. The red here and the red there look different, more so by daylight than by artificial light, and more so when looked at obliquely. The red of the square traversed by green stripes looks almost or quite orange. But we say: nevertheless the two reds are identical, as you will see if you cover up the traversing stripes or look along the length of the stripes. Of course, the red is all one pigment.

Now why do we say the reds are the same? Colour is not the same thing as pigment, so the identity of the ink is irrelevant. Our question is a phenomenological one. Colour is certainly always in some sense an appearance-concept. And my judgement that the red is all one pigment is in fact merely inferential.

I suggest that there isn't exactly a reason, a justification. It is rather that our language just does go like that. Colours that keep on looking the same to the same eye, against the same backgrounds, and in the same light and orientation, *are* the same. That is to say, we regard the colour that things 'are' as not changed by these changes.

If we thought in a particular case that there was a 'real' change of colour when the background changed, that would mean that we supposed something else had happened to affect the colour. Thus the doctor who asks his assistant to check his reading of the blood sample assumes that no such thing has happened. But let us be quite clear about this: the assumption I am speaking of is, e.g., an assumption that the blood is still fresh enough to look the same. It is not, what we might call a 'metaphysical' assumption, that mere change of observer does not *eo ipso* make a change. If some new observer affects the observed – say by breathing on it with garlicky breath – that would have to be something discoverable about him.

Appearing involves a subject that the appearance is to; hence, if there is a concept of 'standing colour', the colour that something *is* as opposed to the colour it only *appears* to be in such-and-such a light or against such-and-such a background, this may seem after all not to be an appearance-concept. For we say: this thing *is* that colour – without reference to an observer. Nevertheless, even this is an appearance-concept, though of a more complicated sort. In the case of our red squares, for example, it is the appearance under certain conditions – with the traversing stripes covered up, or when seen along the stripes – that tells us the standing colour.

If someone should ask why that appearance tells us the standing colour, the answer is that this is an explanation of "standing colour". What it tells us precisely *is* the standing colour – that is the account we must give of 'standing colour'. This is why 'standing colour' is after all an appearance concept, the concept of a sensory quality, even though it applies when there is no sensing going on and is something in respect of which sensation can be mistaken.

What about the other sense modalities? Again, my question relates only to the proper sensibles, the qualities each of which is the object of just one external sense: hard and soft, hot and cold, rough and smooth, sticky and slippery, the various tastes and smells, and the qualities of sound. There are in all of these comparable possibilities of contrast effects. But in the case of sound we cannot speak of 'standing qualities', because sound is essentially temporal. (Kant would have had a better right to call time the form of

auditory sense than of inner sense – for there is no such thing as 'inner sense'.) We make an assumption, of course, that a tuning fork usually makes the same sound when struck; but one is hearing a physically different event. Contrast the (also falsifiable) assumption that a taste pervades the liquid in a vessel. One is of course tasting physically different parts of the stuff at successive trials, or with trials by different tasters. But if a note is an event, then the object of hearing itself is determined as different on successive tests with the tuning fork, whereas the taste, conceived as a quality of the liquid, is not in the same way determined as a physically different thing by the difference of the mouthfuls tasted. Still, there is an extra assumption here. It is a bit as if, in looking at the blood sample and the colours on the scale, each person could see only one bit of the blood in the tube and of the colour sample on the scale. I am recommending that, where Φ is a sensory quality, the Φ of a physically different part be not called *eo ipso* a physically different Φ. This accords with the scholastic observation about qualities uniformly pervading an extension, that they are 'whole in the whole and whole in every part'.

The topic that I have touched on here – of the contrast between sounds and other sensory qualities – deserves much more investigation, and I realize that my own view needs further argument. For all qualities have duration, and so I have perhaps not sufficiently made out what I call the 'essentially temporal' character of sound, as making a contrast with other sensory qualities. It is no doubt also connected with the way in which sound is not exactly a quality of a sounding body. But this is not the place for such an investigation, and I must leave it.

I began by imagining the complaint that the doctor's procedure is 'subjective'. We are now in a position to sort out a number of distinct meanings for "subjective", and hence of its correlative term, "objective". Reflection on these terms should, I think, lead one to vow never to use either of them without saying what one means by it. In using "subjective" as a term of abuse, people often mean "perhaps unreliable; that is, after all, only how it struck him, and we can't rely on that: we had better find out, if we can, by better means". This may be connected with, or amplifiable as, the dismissal of an observation as not guaranteeably repeatable on the same material; too much a function of the particular but unascertainable psychological conditions. This is very likely what would be meant by suggesting that the procedure of reading blood samples is 'somewhat subjective'; someone saying this will perhaps think of blood counts as being more objective. There is a good deal of stupidity mixed into this idea, as comes out in the fact that the procedure is useful. Blood counts give more exact and detailed information; but roughness and unreliability are not the same thing. What lurks behind the accusation, as I tried to bring out, is a suggestion which is false in this case: namely that there is no room for checking for correctness, because what is being judged is the content of an impression. There is a sense of "subjective" in which the word only means that. The changes in the aspects

of jumping figures are subjective in that sense. People experiencing these changes would not dispute about which way it looked now or ask one another for checks. It is a mistake to say that things look yellow to the jaundiced, but I understand that there is a drug, santonin, which does make one see things yellow. A dispute would be absurd between two people who had taken that drug, about which object looked a darker yellow. This, then, is one meaning for "subjective", and in that sense the doctor's judgement is not subjective. His findings do not concern something only accessible to him. But it is found difficult to understand how this is so, because his findings concern a purely sensory object; in saying that the colours match he is giving expression to a sensory impression and the experiencing subject is conceived to enter essentially into such sensory impressions – so that two people cannot be judging the same thing.

Let us try to list some of the senses of "subjective".

(1) We may call a judgement (or impression) subjective because there is no clear idea of a way, or perhaps merely no clear way available, of determining whether the judgement or impression is right. If, however, there is a clear idea of what it is for the thing judged to be true, independently of the judgement, this sense of "subjective" goes over into "unreliable", "useless".

(2) We may call subjective certain inter-subjectively disputable optical objects such as mirages, reflections, rainbows, irridescent colours, apparent sizes and shapes. For example, each man has his own rainbow, which walks with him; yet I might get a man to see that 'the rainbow' does come down in front of a certain bush, from where we are, and feel sure that he has merely failed to notice if he has not seen that we have a double rainbow.

(3) We may call a phenomenon subjective when we mean that the expression of it is in principle not corrigible or disputable, except perhaps by claiming a slip of the tongue – that one was using, but not meaning, the wrong word – or a lie. Claims to feel pain are of this nature, and so are claims to see double, and retreats to "That's how it looks (feels, sounds, smells, tastes) to *me*".

(4) We may call a judgement subjective purely because it is a narrowly sensory judgement. That is what it comes to, when our doctors' procedure is called subjective. Similarly Cavendish's judgements of electric currents by the feel of them, or the judgements by feel of degree of tackiness (stickiness) which people made purely from practice, would be called subjective.

Now this is a philosophical usage. It is assumed that the first three senses of "subjective" really apply in these cases too. Sense (1) does, it is thought, because we merely agree by convention to accept judgements of this sort on the part of such 'competent' people as agree. Sense (2) does, for the colours and colour-matches that each man sees in the blood sample are 'his own' no less than the mirage or the rainbow. Sense (3) does because even if we have a convention whereby we call a man wrong, his impression false, he could not be wrong about the real primary matter of his judgement, which was the content of his experience itself. It is true that the experienced agree, and that

something is done with what is judged by them; that it tends to have coherent consequences. That which is judged then is really inference from the primary judgement, which must be that one has a certain impression. But it appears in the language as the judgement that certain properties are equal, or have a certain degree, or are sufficient for something else. Thus, though the properties are really qualities of sensation, they are not treated as (what they really must be) ones for which *esse* is *percipi*, but as observer-independent properties.

This idea – or complex of ideas – gains expression in the opinion to be found in William James and many others, that the distinction between 'veridical' and 'non-veridical', 'correct' and 'incorrect' is a subsequent, an *ex post facto* distinction based on sorting out the basic material – which consists of impressions, all of equal initial value – into elements of a coherent picture of the world and the remaining elements; these have to be discarded as not fitting in, as 'wild'.

From this point of view the procedure of calling someone else to verify one's own impression requires explanation; in itself it is senseless, so described. But – it may be said – we have constructed a world, in which there are (constructed) objects with what I have been calling "standing colours", and we think our impressions told us of those: we want to know if the other person's impressions will tell him the same.

In the *Theaetetus* Plato describes a phenomenalism according to which no sensory judgement can be wrong, or can contradict another. He rejects this as an account of knowledge, but he seems to tolerate it as an account of what Aristotle was to call the 'proper sensibles'. Each case of perceptual judgement of such properties is, he suggests, a product of an encounter between subject and object; and since all subjects are different, and subjects are different from themselves at different times, and objects also are in perpetual flux, the phenomena can't be in conflict. For such qualities each of us carries the *κριτηριον ἐν ἑαυτῳ*: the criterion within himself.

This account of sensory judgement and its subjectivity assimilates our fourth sense of "subjective" to the third one. We have seen, I hope, how strong are the philosophical motives for such an assimilation.

The account assumes that there is such a thing as correctness of the sensory judgement; that such a judgement has a standard of correctness, and so it ought to be judgeable that it meets that standard. In this way the subjectivity is our fourth kind; the judgement is subjective simply in the sense that it is a sensory judgement, a judgement of a proper object of one of the senses, and so involves the sensing subject. It involves the sensing subject because, though the doctor may just say "the reds match" this is as it were an extract from "I see these reds to match". Here there is such a thing as correctness – but it is also possible for his judgement to be wrong. But Plato's account of the sensory judgement tries to hang on to there being correctness in it, while at the same time treating the judgement as incorrigible and uncheckable and incapable of getting into conflict with any other one. The judge 'has the criterion within himself', and this apparently guarantees his not going wrong.

Now if this means that the judger has some distinct standard 'within', and according to that standard he judges that he has rightly judged that something is so, then we must ask why the possession of a standard should guarantee that he can't go wrong in applying it. A real distinct internal standard is quite possible; I might call to mind the colour of someone's hair that I know in judging whether to call someone else a "honey blond". Or I may mentally place myself in some well-known environment, where I know that if I face up this street, or the like, I am facing north, and that then such-and-such a building, which is on my right, lies to the east – in order to orient myself, not having an immediate grasp of the relations of points of the compass. In such cases the possession of a standard or criterion is insufficient to guarantee correctness. There is no objection to it, but it does not do what Plato seems to think is done by the 'criterion within'.

The suggestion is not really of a distinct standard. It is that the sensation itself provides the criterion by which its owner's description of it is assured of truth.

So the view comes to this: a narrowly sensory judgement manifests or expresses a sensory experience, and in it – perhaps, but not necessarily – something is judged to be qualified by a proper sensible. (For in this way of looking at things, pains are naturally assimilated to secondary qualities.) Every such judgement either is or contains a judgement in which the sensation-content is identified and judged to occur. The sensation itself supplies the criterion by which this latter judgement is assured of correctness.

That is to say, having the sensation itself gives one a way of telling that that judgement is true, and such a way of telling that it is certain to determine that it is true. But how is that? We can say "Here are Koplik's spots, therefore here is measles" but what is the use of saying "Here is a sensation of red, therefore here is a sensation of red" or "Here is a headache, therefore here is a headache"? If we frame a statement of the occurrence of the supposed criterion for the sensation-statements we shall just be framing the sensation-statements themselves.

The inclination is to say "My headache tells me that I have a headache – and infallibly". But how can that be? If it can tell me something, why may it not tell me falsely? At this point one wants to say: it tells me infallibly that I have a headache because I can use it as a paradigm of what I mean by having a headache. We have already said something like that: "If I meant blue when I said 'blue', then could I not also have given the colour I was calling 'blue' as a paradigm of blue – of what I meant by 'blue' – at least to myself?" Thus my expression is verified by something I could use to give it its meaning.

But what do I mean by "its meaning"? The meaning it already has? If so, then, I have got to be right, I have actually got to mean blue when I say "blue". I cannot determine that I do so just by concentrating on the impres-

sion I have when I judge that something is, or looks, blue. I might say "The name '*P*' that I give in introducing the man *P* gets its verification from the very object (the man *P*) that I could use to give the name its meaning." (I do indeed so use him if I give his name meaning for you by introducing him to you as *P*.) But I could not determine that by "*P*" I really mean *P* by concentrating on the man.

Of course I might decide now that (whatever has been true in the past) I am now going to mean this man by "*P*". In that case there is no verification in question – until I want to speak of him, to identify him, again. If I do this in the case of a man, I shall rely on the usual standards of identity for being justified in using the name "*P*" again and claiming to mean the same.

But suppose that I decide now that (whatever has been true in the past) I am now going to mean *this* colour by "blue". In that case, again, there is no verification in question – until I want to speak of it again. But can I say: if I do this in the case of the colour I have an impression of, I shall rely on the usual standards of identity for being justified in using the name "blue" again and claiming to mean the same? What 'usual standards'? There are standards, no doubt, for objects being the same colour. But I was speaking of the colour of an impression, and here there are no standards, only an inclination to say 'it's the same'.

The idea of the 'criterion within' is, I think, the same idea as the one that Wittgenstein attacked under the title of the 'private ostensive definition'. At least, it is so in the final form, in which we are supposing I can take a sensation I have now, decide that I mean this by a sign, and so assure myself how to use the sign. There are an important few lines of argument in Wittgenstein which are so succinct that they have not been understood as argument – indeed, it has even been denied that they are an argument.

> I impress on myself the connection between the sign "*S*" and the sensation – But 'I impress it on myself' can only mean: this process brings it about that I remember the connection *right* in the future. But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'. (*Philosophical Investigations*, I, 258)

We shall understand the argument best by resisting it step by step.

(1) "'I impress it on myself' can only mean this process brings it about that I remember the connection *right* in the future." No, it means that an impress or copy of the sensation remains in my mind with the sign annexed to it. – But a copy can be good or bad, true or false. 'I impress itself on myself' won't be relevant unless it secures that I use the sign only in its connection with a good copy, and that means my remembering the connection of sign and sensation right.

(2) "But in the present case I have no criterion of correctness." We object: I have a criterion in the form of the impress or copy. But I must judge of it "This is *S*", and that is already to make a new use of "*S*". A correct one, or

not? The machinery I have assumed to explain how I can make a correct new use of "*S*" already invokes a capacity on my part to make a correct new use. So that, after all, cannot explain how I can make a correct new use. Hence we must give up the idea that I use "*S*" in appealing to my former sensation.

But perhaps I simply recall my former sensation and call "*S*" whatever is the same as it. I don't say "That was *S*" but rather "This is the same as that, which I called '*S*', so this is *S*". But how do I know I have got hold of the right '*same*'? To show something and say "Whatever is the same as this, I call 'Jack'" is not yet to explain what I will call 'Jack', for it is not yet clear what is to count as the same as this. If "*S*" needs definition, to define it requires fixing the kind of identity with the paradigm which is to justify using "*S*" again, and merely attending to (or recalling) what I called "*S*" will not give me the kind of identity, if I don't already know it.

"But," I may reply, "it can certainly seem right to me, when I say "*S*" again; and I mean it in a particular way, so I must somehow have got hold of some idea of the same, in order to want to use "*S*" again at all, with the feeling of using it in the same sense." – Is that enough? Then, if so,

(3) "Whatever seems right to me will be right" – we have come to this thought of ourselves. At least, if I keep on thinking I am right. But we can still make a stand against the last move:

(4) "That means we can't talk about 'right' here". For there can be two reasons why whatever seems right will be right. One is, that it doesn't matter what seems right, so long as it seems right anything will do. Admittedly in that case we can't talk about 'right'. But the other reason would be that only what is right can or ever does seem right.

But how can that be? To say that only what is the same as before (in some particular form of sameness) can or ever does strike me as the same is certainly false wherever we can give an account of being the same which is independent of seeming the same. Here we can give no independent account, so if we say "It's got to be what is the same that seems the same", then either we are believing something against all reason, or the seeming is really all that matters. And if so, then we can't talk about 'right'.

We see very clearly that what we really want of the private ostensive definition or criterion within is a justification of the feeling of certainty about our (narrow) sensation statements: we are really happiest with the idea that the sensation itself offers the criterion for the truth of the statement that we have such a sensation: each time I use a word for a sensation as I have it, the experience itself guarantees both meaning and truth. We don't really want to cast back to any previous use of the sensation-word. For if pressed we must admit that the certainty of meaning the same as before, or having what one has had before, must dissolve under Cartesian doubt.

It is wrong at this point to object to Cartesian doubt. For the sort of certainty that we are dealing in here is precisely Cartesian certainty. That *this*, which I have now, is what I (now at least) mean by "warmth" or "blue" or

"pain", and that I can't be wrong about it – that is the most natural thought in the world, and is what inspires the Protagorean phenomenalism that Plato describes. But we have seen that if it is a matter of the relation of the name of my present sensation to my present sensation, if it is guaranteed as the name of that sensation because that is what I now mean by it, no verification or truth is in question after all.

And there is something right about this. If I suffer from 'red vision' (which is subjective in sense (3)), when I say "I see the red colour suffusing light-coloured things" I do not mean "I see a colour which is the same as I have previously called 'red'": If someone else wants to know what colour I mean, he will check that I use "red" as a word for objects visible to me in a normal light (and whose colours are not previously known to me) which are red. But I don't make any such check on myself. If I discovered, however, that my colour vision had become abnormal, so that I wanted to call objects "red" which no one else did, that would not raise a doubt in my mind about whether I was ('subjectively') seeing red. At most, I might lose my sense of meaning anything by "red", no longer have facility in the spontaneous use of the word "red".

If what I have been arguing is correct, then there follows the important conclusion that we ought not to say "I know . . ." in connection with statements of sensation. The suggestion I noticed earlier is right: contrary to Cartesian assumptions, the 'incorrigibility' of descriptions of sense-contents (narrowly understood) is not any kind of correctness. The correctness of sense impressions consists in the correctness of judgements about something else, which is made when we have them: about standing colours, for example.

There remains a certain conception of 'subjectivity' to which I have made a passing reference. I spoke of the way 'the subject' is thought to enter essentially into the sensory. The doctor, we said, may say "The reds match", but the explicit statement of the case is "I see these reds to match". If so, then a Platonic argument may force it on us that the 'objective', in the sense of what is the case regardless of an observer, is never sensed, but is always inference or construction from the sensorily given cues. What is given is always a product of the encounter of subject and object and so properly belongs to the encounter and the moment at which it occurs. This would be destructive of the idea of observation. If something is a correct observation, we ought to be able to detach what is observed from the statement of the observer's perception of it, and simply say it was so.

Now it is perfectly true to say that the explicit statement of the case is "I see these reds to match", and if "I" is a name of something involved in this, it is difficult to see how the detachment could ever be justified.

We might take the following line: "I" is a name of the person having the sensory experiences; it is the name that *he* uses for that person. The 'subject' that is always involved is the human being who gets the impressions giving information about, e.g., the match of the blood sample, and it is an empirical question whether his sensory judgements give reliable information.

This conception of the first person singular is mistaken. But it will be useful to give it a run for its money. I will invent a character, which I will call "an insensitive logician", who shall express this view as follows:

> When there is sensation, of course there is one who has it. A report of something sensibly perceived, therefore, needs to give not only such-and-such a content, it must say that it was presented to someone. Otherwise *either* we shall have only the content given (perhaps with time and place added, e.g. electric current of such-and-such strength in wire A at time t) and then it won't be clear that this is actually a sensory report and we won't know why we should believe it; *or* there will be an obviously incomplete statement, e.g. "Sensation of electric current in wire A at time t". But how the sensing subject is given is indifferent to what is said. If the report of sensory experience is given by the one who has it, he can speak in the first person. If by other people, they will give some other designation of the same subject. Exactly the same will be reported in either case: a grammatical subject will be given, and an attribution of sensory experience made in respect of that subject.

The first thing to notice about this view is that "we won't know why we should believe it" would apply to the third party's statement that someone had this sensory experience. But if this view of the first person singular is right, it becomes puzzling why the first person report should be in a different position.

The truth is, however, that "I" is not a name at all. Of course, everyone will agree that it is not an ordinary sort of proper name, it could not have the characteristic uses of a proper name – but this is a very trivial fact. It does not determine whether, logically speaking, "I" is a proper name, albeit of a rather special sort. And this, of course, is what my imaginary logician is maintaining.

It is useless to explain "I" as a name which each person uses of himself. This is readily seen if we consider the proposition "*NN* does not realize the identity of the person he calls *NN* with himself". In such a proposition, if we are to understand the reflexive pronoun in the standard way, we have left unspecified the identity that *NN* is said not to realize. We have specified for ourselves the object we are saying *NN* fails to realize something about: that is indicated by the subject of our sentence, "*NN*". But we have not indicated how *NN* would conceive the object, if he did realize the identity. In fact, of course, we know this very well. We know what we mean he doesn't realize: it is "I am *NN*". But in that case we have not explained "I" at all by giving the above supposed explanation. And this in turn shows that the sense in which "I" is not a proper name is not the trivial one of which we first thought.

We may put it like this: there is no path from "I" to the person whereby he connects it with an object (the person that he is) which it names. This is the principal root of the philosophic idea of 'the subject' – that "I" does not stand for any object, not for anything presented. Or, as Berkeley put it, there is no 'idea' of the self.

The point comes out again in the absurdity of the question whether one

has made attribution to the right thing when the 'attribute' is sensation and one says "I . . ." as the subject.

Nevertheless, if one speaks of 'the subject' in that philosophic usage, it is difficult to avoid the idea that one has introduced an extraordinary sort of object: something that is indeed not presented, because it is what presentations are made *to*. It is not the human being, nor yet his soul, but something of a quite different kind, which is always involved in thought and in sensation and perceptual judgement. If so, then sensation is in an important philosophic sense subjective; 'the subject' must as it were be inevitably involved in the sensible properties, it cannot but infect them; and any reaching out beyond this contaminated given must be illusory, and its object mere construction.

While we must reject the 'insensitive logician's' view of "I", nevertheless this opposite one is no better. The essentially first-person character of the sensory report must be granted, but it does not introduce any such thing as 'the subject' is conceived to be. It does not introduce any *thing* at all, precisely because "I" is not any kind of name.

5 Events in the Mind

I *Mainly Problematic*

"Just as I was getting out of my chair to fetch the dictionary the clock struck eight and I suddenly remembered that I hadn't yet sent the Christmas card I meant to send to John." I suppose that this could be as good a report as one could want, of an event in the mind. The character of what it relates, as an event, is secured by the temporal reference: "Just as I was getting out of my chair . . . ." Suppose the sentence comes in an account of some few minutes in the narrator's life – he might be explaining how he came to be out on an evening when he had said he would be in, working. A child could understand it. Yet for a philosopher it is full of problems. The ones I can think of are these:

(1) Is the sentence "Just as I was getting out of my chair . . . I remembered" an example of loose speech? It certainly sounds as if it were the same subject that got out of the chair and that remembered. But getting out of a chair is physical and remembering mental or spiritual. Perhaps if one were really exact one would try something like the following restatement: "This body got out of the chair it was sitting in while I had the intention that it should fetch the dictionary, and then I suddenly remembered that this body had not sent the Christmas card which I had the intention it should send so as to reach John." What if anything is right about this? And, if one is going to produce a contorted version, why not "I got out of the chair I was sitting in while the mind attached to me intended . . ."

(2) Suppose the narrator somehow took in what he now recalls, how did he take it all in? It is a familiar argument that if you hear a bang and see a flash, you don't *hear* or *see* the simultaneity because you don't see the bang or hear the flash: so, it is argued, there must be something else, besides sight and hearing, that takes in the seeing and the hearing. Extending this argument, in our case we can ask about the knowledge of the simultaneous intentions and memory and physical movement as well as the sight and sound of the clock. At least our narrator (a) knew three things: (i) that he was getting out of a chair, (ii) that the clock struck and (iii) that these were simultaneous, and (b) had (i) the intention of getting the dictionary, (ii) the thought of another intention and (iii) the thought that he had not carried out that intention. His later memory was able to put all these together as simultaneous. Are we to suppose some faculty operating at the time which took them all in and noted their simultaneity, as one's hearing takes in two simultaneous sounds?

(3) But what is an intention? Can one speak at all of *taking it in* that one

Unpublished paper presented to the Priests' Philosophical Group, London (1963).

intends such and such; or again of *taking in* the fact that one is remembering that one hasn't sent a Christmas card?

(4) Suppose we say that intentions and memories are a sort of thoughts, expressible in these cases by, say: "I'll get the dictionary" and "Oh heavens! I meant to send a card to John and haven't". In the case of either thought, it *might* have been that he said the corresponding words or something like them, out loud or inwardly. Then the words actually uttered or said inwardly will have been the vehicle of the thought. But suppose neither they nor any other words occurred, but something occurred which was, say, the sudden memory, what then is the relation between what occurred and the words that you would produce in order to say what the memory was of?

Sometimes one wants to say that the words, say "Oh heavens! I've not yet sent that card to John", *give the meaning* of what occurred inwardly, but at other times one wants to say that what occurred inwardly *is* the meaning of the words.

The first inclination arises as follows. If you were the narrator and were asked: "What can you remember as having gone on, what, that is, did you *experience* when you either 'were having the intention of getting the dictionary' or 'were having the sudden memory of not having sent off the Christmas card'?" it may easily happen that your memory shows you nothing adequate. I mean that you may remember looking at a word whose meaning you couldn't think of, and knitting your brows over it, and then after a short period, in which nothing useful occurred to you about it, you got out of your chair; and you say with absolute confidence that *if* anyone had asked you what you were going to do the right answer would have been "to get the dictionary". But you may not be able to point to anything that determines that that is the right answer: no explicit thought "I'll get the dictionary", no visual image of the dictionary somehow presenting itself as an object drawing you to it. You did not in fact get the dictionary because at that moment the clock started striking and the thought about the Christmas card led to your going out. But then, in its turn, the thought about the Christmas card may be difficult to fix as an item of experience. Suppose, as is likely enough, you did *not* say all those words: "Oh heavens . . ." Say you heard and looked at the clock, and, *because* the last post goes just after eight, something clicked in your mind and you hastily rummaged in your desk, found the card, addressed an envelope, went out and posted the card. That's the best account you can give, if asked what went on *in that* you had that sudden memory. Now the click in your mind was, we might say, like a small jotting in a notebook which *means* something quite lengthy. Thus the words *give the meaning of* what occurs inwardly.

The opposite inclination, namely to say that the inward thing *is* the meaning, arises when one reflects that words are noises and marks and how it is only because they express thoughts that generally speaking they are words at all. Thus suppose the narrator had actually said "Oh heavens! I've not sent that card," he would have something at first sight very adequate to

report in answer to the question "What went on in that you remembered you hadn't sent a card to John?" He could say: "I actually exclaimed 'Oh heavens! I haven't sent that card'." But then, what do the words *relate* to? Only the thought they express shows that – shows, e.g., which card "that card" was. The words don't after all determine the thought. He could have said "Oh heavens, that card!" or "Oh heavens!" alone, with just the same thought. He might just have had a startled feeling – the *thought* is the meaning of whatever did occur.

When asked: "What does your memory show you as having gone on *in that* you suddenly remembered?" is it fair to restrict one to *experiences* like words and images and feelings? Why can't one reply: *a gasp which meant "Oh heavens I haven't yet sent that card I meant to send to John"*? Or even, may it not be that there was absolutely nothing which occurred in one and meant "Oh heavens I haven't yet sent that card", but that one simply had the thought in a flash, which would be expressed by those words, and that is what one remembers: it's no use asking what it consisted in because it didn't consist in anything or have any vehicle, it was only the thought itself. Yet, deprived of a vehicle, of the slightest flicker which should signify it, the naked thought seems like a nothing when one tries to recall what it was – which in another sense one can say perfectly well: one simply gives the words which express it.

(5) Here I have been concentrating on the thought which was a sudden memory. If I return to intention, it can seem much less substantial even than the naked thought of something one had not yet done. For, once more, what is an intention? If it is a sort of thought which occurs in a man's mind, the same problems of course arise for, say, "I'll get the dictionary" as we've been considering for "I've not yet sent the card". Of course, a man could hardly simultaneously *say*, even inwardly, "I'll get the dictionary" and "I forgot to send the card". But it is clear that he can have the thought "I forgot to send the card" while he is getting up to get the dictionary, and hence while he is intending to get the dictionary. Now does this mean that he can at least have the two *thoughts* "I'll get the dictionary" and "I forgot the card" simultaneously? But how can he *remember* that he means to do something which he has not yet done, if an intention is a thought and a thought is something present to consciousness at the time when one has it? An intention after all needn't be a thought, for one can intend what one is not thinking of, as when one intends over a whole period to make a certain journey, but in fact seldom thinks of it, and when one even thinks of it, one's thoughts aren't to the effect that one is going to make that journey. Some writers call this *virtual* intention, as if an intention when you are *thinking* that you will do such and such is somehow more actually intention than intention when you are not thinking of what you intend. But is that really so? I know no good reason to think it. We tend to think it out of a prejudice that an intention *must* be a mental phenomenon, i.e. an event in the mind.

(6) "I suddenly remembered I hadn't sent John's card off." Is the identification of John part of the thought? Say I know several people called John.

I am only thinking of one of them. But am I *thinking of* whatever would distinguish the one I am thinking of from the others? Suppose he is my cousin and a farmer in Wales, and that those points would be both enough to identify him, and what I should say to someone who asked me which John I had this sudden thought about. That does not mean that I then thought either of the cousinship or of the farm or of Wales. This suggests that a thought may be something that *has* meaning, rather than itself being the meaning of words.

(7) A sudden memory might seem plausibly to be a certain experience – say the having of a mental picture of what has happened. But the example was a sudden memory that something had *not* happened. You can't have a picture of something's *not* having happened, so here at any rate the thought, which is the memory, is not a mental picture, or memory image, even if its vehicle is such an image – a vivid mental picture of the chosen Christmas card for example.

II *Historical*

Descartes enlarged the application of the notion of "cogitare" or "penser" so that even a toothache is for him a *cogitatio* – so long as "toothache" is so understood that one can have toothache without having any teeth; or, to take a more extreme (but actual) case, one can have pain in a limb that isn't there: pain, dissociated from any physical reality (because physical realities cannot resist the corrosion of the method of doubt), is a *cogitatio*. The notion of *cogitatio*, that is, has turned into that of *consciousness* or *experience*. Cogitationes, Descartes says, comprise sensations as well as acts of understanding (i.e., I believe, thought in the restricted sense of "a thought of such-and-such"), will and imagination: everything that takes place within ourselves so that we are aware of it, precisely *qua* object of consciousness.

By making this move Descartes assimilated all members of a class known to present-day English philosophy as "psychological verbs". This is a genuine classification: there is one important feature common to all psychological verbs: this feature was called by Wittgenstein the "asymmetry" of the first and third persons in the present indicative. If I say I have the beginnings of multiple sclerosis, or a wart on my wrist, then that should be for the same kind of reasons as lead me to say someone *else* has the beginnings of multiple sclerosis, or a wart on his wrist. It is a matter of observation. But if I say I have a headache or am thinking about spiders or want an apple or now see the Crystal Palace in my imagination, or that I believe someone's story, that similarity to the third person vanishes. It may be senseless to speak of my having reasons, grounds, for what I say, or it may be good sense. It is senseless in the case of "I have a headache" and good sense in the case of "I believe his story". Senseless in the case of "I have a headache", for what could the ground be? The headache? The datum itself? How can a *sensation imply* anything? It would have to be the *statement* of sensation that was the

ground: but that is simply "I have a headache", which was the statement in question. It is good sense in the case of "I believe his story": but then one's grounds are grounds for believing, i.e. grounds tending to show that the thing believed is true. In either case, there is a *dissimilarity* between the first and the third person in the simple (assertoric) present indicative: consider the *grounds* one might have for "*he's* got a headache" and "*he* believes that story". Equally, though I may have grounds when I say "I want an apple", these would be, generally, not reasons for *judging* that I want an apple, but reasons for wanting an apple, whereas grounds for "*He* wants an apple" should be reasons for *judging* that he wants an apple.

This asymmetry justifies us in speaking of the verbs that display it as a single class. The facts about the use of the first person present indicative also show why it is more natural to say "While this body got up *I* remembered" than "While I got up, the mind connected with me remembered".

Descartes introduced the word "idea" to express what is immediately perceived by the mind, and he sometimes writes as if confining it to *representations* of something or other, either in the intellect or in the sensation – i.e. he so explained "idea" that wanting or intending were not themselves ideas. But when pressed by Hobbes he gave up that restriction, and called volitions and intentions 'ideas' too – thus assimilating "I have an idea" to "I have a cogitatio" and treating *all* as if they were contents of experiencing, as if "I intend" were the report of an experience–content called an "intention", as "I see blue" is a report of an experience–content called a "sensation".

III *Attempt at a Solution-in-Chief*

Let us ask the curious question: Is a thought, in the sense in which we may speak of a *sudden* thought, really an experience, a cogitatio in the sense that "cogitatio" has now assumed? Well, what is the expression of a thought? Say it is: "Letters have to be posted by 8.15 to catch the last post." *That* is *not* a description of an inner experience: it is a proposition about the posts. The expression of an inner experience might be something like a cry of pain or the statement "I've got a toothache". The proposition "Letters have to be posted by 8.15 to catch the last post" is not in *that* way an expression of a thought. One would say rather that it is, or perhaps that it contains, the thought: asked what the thought was, one gives the proposition as an answer and it is not a communication giving the character of an inner event or state but one giving information (right or wrong) about the posts.

But *isn't* it a communication about an inner event? It is as if the right expression for the occurrence of the inner event would have been: "A thought!" as one might say "An image!" or "Pain!" or "Colour!" I take these are different typical examples of experiences or contents of experience: for the notions of an experience and a content of experience go together. Looked at in this way, the words telling what the thought was are a *closer* specification of the experience (like saying that the colour in question was

red) or a description of its content (like saying that the image was an image of Kennedy pressing the button for war).

But – to repeat – the words telling what the thought was are *not* that: they are, e.g., a statement about the posts.

But does a thought really differ from an image in this regard? Is not "Kennedy pressing the button" about Kennedy and button-pressing? What is the difference between "I had the thought 'suppose Kennedy pressed the button'", and "I had an image of Kennedy pressing the button"?

The account of the thought contained the proposition "Kennedy pressed the button". *That* is not a description of inner experience.

One is very strongly inclined to think that, because "I had a thought *then*" or "I suddenly remembered" refer to particular moments of time, the right thing to look at is something that took place then, as you would look at the mechanism of a clock when the clock struck to see what happened then, for the clock striking is an event in the history of the clock mechanism. But the foregoing considerations tend to show that that inclination is a mistake. For if we want to know what the thought was, we have to consider something that need *not* have occurred when the thought occurred, namely the words and the explanation of the words, which give the thought.

Yet, once more, isn't the same true of the image, which I allow to be an experience? I mean: if I say "I had a sudden vivid image" and am then asked "Of what?", I reply, say "Of Kennedy pressing the button for war" – don't these words relate to Kennedy and button-pressing and war?

And in the same way, suppose I am asked "Why did you blink?" and I say "Because I saw a flash of bright red just before my eyes" – isn't the word "red" the name of the colour of, e.g., English pillar-boxes? A visual experience of a red flash is certainly something I should call an experience with a content.

I think that the answer to this is that the language that describes images, or again sensations of the exterior senses, is there having a secondary application. "I see a snowy landscape"; "I see a snowy landscape in my mind's eye". When I say the first, believing that there is indeed a snowy landscape there which I am seeing, the words "a snowy landscape" have their primary application. When I say the second, describing an image, they have a secondary application. They do not have the same consequences, for example. Such questions as "How deep is the snow?" "How long has it been there?" "How soon will it melt?" must be admitted to arise, even if one does not know the answer, when the snowy landscape is supposed to be real; but *may* be waved aside as not coming into the picture, when I see a snowy landscape in my mind's eye. The same would hold if I were knowingly hallucinated, so that I saw a snowy landscape where in fact there was nothing but the interior of a room before my eyes, but I knew that the snowy landscape was not there.

But the language giving a *thought* is having its primary application. "John was here last week, I saw him" – that is a likely expression of memory which I have *now*, but it is a description of something past, not present. It is still that

when I prefix it with: "What you say suddenly calls to my mind that . . ."

The propositions which perhaps occur in accounts of thoughts are not necessarily ones stating things that were thought *true* (or false). In several of my examples the propositions have been the thoughts and the thoughts have been thoughts *that* such and such was so. But in one, the thought was given by "If Kennedy pressed the button . . ." A proposition occurring in an "if" clause is not asserted or denied. But the words are still having their primary application. One thing that shows this is that one can say: "Then, what?" – the reply being, perhaps, "Then luckily someone interfered with the connections between Kennedy and the thermonuclear missiles."

An image cannot in this sense have consequences, be the protasis for an apodosis. The words "image" and "imagine" are indeed connected, through the connection of each with the word "imagination". On the one hand, the imagination is often spoken of as if it were the medium in which images occur, and on the other to imagine is to exercise one's imagination. It is important to realize that imagining may well be done without images. Imagine the injunction "Imagine a door which is bigger on the outside than on the inside". Some people may have an image which helps, and some, images which are of no use, and some may have no relevant image. But those who have images have not carried out the injunction, unless they can discuss the matter and show that they have understood the description; and someone who does that shows that he has carried out the injunction, even if he has had no image. Such imagining is thinking. The injunction might as well have been "Think of the following". My remarks about images as experiences with contents have not been concerned with this sense of "imagination".

Three different kinds of mental report have emerged: first, there are reports of experiences – sensations and images – which may be called occurrences and which have contents. Second, there are reports of intention, understanding, knowledge and belief; it is characteristic of these that though one may have intended, etc., something at a certain time, that does not require that what one intended, etc. at that time be before one's mind at that time. Here intention, understanding, knowledge and belief are not events; nor *need* they be started off by a forming of intention, an occurrence of insight, an occasion of learning or an initial judgement. Finally, there is thought *of*, or *that*, such and such. This is an event, but not an experience; what we call its content is given by words which do not describe an inner experience (unless that is what the thought was about) but which have their primary application.

My main conclusion is the engaging one that one thing that is not a "thought" in the sense Descartes gave to the Latin and French for "thought", is: a thought.

6 Comments on Professor R. L. Gregory's Paper on Perception

A theory is presented in the face of a problem or set of problems. I think there is one prime problem in the face of which Professor Gregory presents his theory:

(1) We perceive objects. Using 'O' as a variable, we may say: an object, an O, is an O only if it has various properties not given in sense at the time of the perception of an O. Yet what is perceived is perceived *as* an O. (Possible values for O are: a box, a cat, a rainbow, a whirlpool, a sentence.)

This problem has often led philosophers to draw a distinction between sensation and perception. For example, Russell says:

> A certain experience E (e.g. that which is the visual core in what we call "seeing a cat") has in my previous history been usually[1] closely accompanied by certain other experiences. Hence, by virtue of the law of habit, the experience E is now accompanied by what Hume would call "ideas" but what I should prefer to call "expectations".[2]

Again, on the same page, Russell speaks of "the complete experience consisting of a sensory core supplemented by expectations"; this he calls a "perceptive experience", not a "perception", "because the word 'perception' suggests too strongly that the beliefs involved are true". In fact, Hume's term "ideas" is superior to "expectations" because this *perceptive* experience which has to be distinguished from the *sensory core* that is part of it, may not involve actual beliefs or expectations. This comes out in several of Professor Gregory's examples. (I cite Russell merely as a fairly typical example of a philosopher addressing himself to Gregory's prime problem.)

We have to be able to say what the 'sensory core' is, if we are to draw this distinction. And the trouble about this is that, in showing how perception goes beyond the given, one has fluctuating accounts of what is given or sensed. Or at least the examples are of varying status. For example, in Gregory's own exposition, we have 'seeing three legs of a table' and acting as though it had four, and also sensing a brown patch and acting as though it is wood.

I am sure this was deliberate and I don't think it can be avoided. Locke and Berkeley indeed thought that the visually given was a distribution of

[1] Note that in the non-trivial sense intended by Russell this is false. I have seen cats *far* more often than I have touched or heard them, and I am sure the same was true of Russell.

[2] *An Inquiry into Meaning and Truth* (London, 1940), p. 121.

Paper given to a conference at the University of Kent sponsored by the Royal Institute of Philosophy, 1971, later published in *Philosophy of Psychology*, edited by S. C. Brown (1972).

colours, light and darkness in a plane, but for inadequate reasons. Locke says it is obvious from painted pictures! Nor does he so much as ask where this plane is. This problem does in a way strike Berkeley: his approach is *via* a rejection of the Cartesian explanation of distance judgements by the angle subtended at the eye. Berkeley says no such angle is seen. This leads him to doubt that the 'outness' of what is seen is *given* at all. He seems to be desperately confused between 'in the mind', 'under the skin', 'behind the eyes', 'on the retina'.

Now I say, and Berkeley too would say, that I see *the angle* made by the line of a bell tent and a distant hillside. I might draw it. Asked to measure it, *I* would hold my protractor vertical and parallel with my forehead as I look towards the tent, eyes front. We give this projection a privileged position. It is perhaps natural for us to do so. By now, at any rate, it would strike us as unnatural to orient the protractor differently. If instead of a protractor we 'naturally' used some queer object like Gregory's structure for the impossible triangle, with an adjustable arm, we might insist that it was very unnatural to use an object in which the arms making the angle were in the same plane. The idea that the pure visually given must be a distribution of colour patches in a plane disintegrates upon consideration. 'The plane' is really a somewhat sophisticated conception, such as was formulated by Leonardo. But as things are for us: isn't *that angle* likely to be counted as sensed? How far this is culturally determined I do not know.

Professor Gregory rejects the philosophical idea of sense-data because sense-data are supposed to be experienced, and it appears that on this point Berkeley was wrong; we go by cues that we are not aware of. Thus Gregory comes down on the side of Leibniz who thought there were unconscious sensations. But he wishes to say that perceptions are hypotheses. Now an hypothesis is something answerable to evidence, to data. To what data could the perceptual hypotheses that Gregory speaks of be answerable, but to perceptions? *Are* these perceptions then in turn hypotheses, and so on *ad infinitum*?

That seems wrong. Are they then answerable after all to sense-data – not indeed to the cues we go by without being aware of them, but to the *conscious* sensorily given, which by a few examples we are easily persuaded to say perception transcends?

For how does one know that perception goes beyond what is given to sense, if one cannot *say* what is given to sense? This bit of the difficulty can be overcome by speaking of the sensory *input*. The same object – a written text, say – will give about the same input to two human beings with normal eyesight. (Note that the condition is necessary: you wouldn't take two people one of whom had diseased retinas and call the input the same.) The images in the retinas will be the same – that's input. Now if I vary the object in ways the subject maybe doesn't notice, but affect what he perceives, I call the property I have varied a property of the input. I now reinterpret the notion 'character given in sense' or 'sensed characteristic' as: property of the input.

This enables *me* to say what is sensorily given to someone *else*, which he cannot do just in virtue of its being sensorily given to him. Perhaps for example he identifies fruit woods; I am an investigator who can find out what cues he unconsciously goes by. The same problem can now be restated:

(1a) Perception at a given time cannot be accounted for just by sensory input at that time.

Yet the difficulty about what the hypotheses are answerable to is not solved by this move. Perhaps the solution is that they are answerable to situations of 'richly redundant' – i.e. adequate – sensory information. But once again, is the perception of the object in *such* a situation supposed to be an hypothesis?

Two different types of case have surely got confused here. In one, what is given to sight is much less than might be, in the other plenty is given. For example, I am looking in a darkened room for a small transistor radio which I am familiar with. I see two faint parallel horizontal gleams, just right to be the gleams of two bits of metal trim on the radio set. I stretch out my hand suitably to grasp the carrying handle if these gleams really are what I suppose, and we will assume that I am confident of having picked the radio out perceptually in the dark of the room; a measure of this may be the confidence with which I go and pick it up. So, suppose I say, "I saw the radio in there", then this perception *is* very like an hypothesis. In this case *I* can point to the sensorily given (though there may also be cues I am not aware of), I can contrast seeing those gleams with the hypothesis "that's the radio", and I can point to the data to which this hypothesis is answerable, namely the full-blown tactual and visual perception of the radio as I carry it into the light.

In the other type of case I am, say, looking at lighted lamps on a table in clear view. The information is already 'richly redundant'. If I want to call my perception "hypothesis" *here*, I must advert to the fact that I see 'only' views of particular parts of the lamps and table. A main point in calling the perception "hypothesis" is that perception goes beyond what is sensorily given. So I must not count lamp and table as sensorily given – but no longer for the reason that held for the radio in the darkened room. The radio was not 'sensorily given' because there was only a very sketchy indication of where it was, whereas this lamp is given to sight as fully as it is possible for it to be: the 'more' that is involved in perception, as opposed to what is given in sense, *could* not be given in sense at one time.

There is thus implicit in Professor Gregory's theory the following thesis: *It is in general impossible that objects should be sensed*. We are hereby plunged into a second problem:

(2) What do we see? We do indeed *say* that we see the lamp (for example) – but the matter is not to be settled by an appeal to what we say. For how can I say that the lamp is sensorily – in this case visually – given? At any rate I can't say the lamp is the input. If we speak about what is given in sense, not explaining this as 'input', then the first problem reminds us that there is much more to a lamp than can be seen at any moment, i.e. I *could* see what I see but it might not be a lamp. Now our question is: what do we see in *that* sense?

If someone should happen to doubt whether this is a real question, cavilling over the assertion, "I could see what I see and it not be a lamp", we can call to our aid the facts about double (and also singular) vision. Attention to these facts should prevent one from philistinely saying: "I see the lamp"; so 'what I see' just is the lamp and there is no excuse for introducing something else, some intermediate 'object of sense'.

When one sees double, Gregory asks, which of the images is veridical? The question itself calls in question the meaning of "veridical". For both images are presentations to the apparatus of sight; each derives from the object. If, as Gregory suggests, I try to establish one by stretching out my hand to see which I must go for in order to touch the object, and one does come out top, this shows nothing but that I am muscularly more co-ordinated with the sight of one eye that with that of the other. The curious thing is that both images are in a unitary visual field. How do I judge that I touch this presentation rather than that one? Perhaps I *see* that I do. But may there not equally be two images of my finger? If so, then ought I not to make the image of my finger that is connected with eye A go to the corresponding image of the object? It may happen that I don't readily do that when I 'try to touch' the presentation of the object to eye A, but rather bring the finger as presented to eye B to the place of the A-image of the object. Then of course the finger does not touch the object. But if I shut eye B, I shall touch the object, bringing the A-image of the finger to the A-image of the object, without difficulty or calculation. Thus at first, because when I have both eyes open I fail to touch the object by going for the A-image, I may be inclined to think "*That* image is not 'veridical'." But on reflection that does not seem a good reason for saying so.

What then is 'non-veridical' in this situation? It is the unitary field with two perceived objects in it: this is non-veridical because there are not two objects. But neither perception separately can be called non-veridical just because of the existence of the other. One image may be blurred or wobbly and so in those respects non-veridical, i.e. not according to the way the object is. But if the two are produced by holding a rod up and seeing past it on either side with both eyes, so that one sees two of, say, a small piece of paper a little way off, they may be perfectly similar and neither superior to the other.

The interesting thing about the case of double vision is the way in which it compels us to speak no longer of simply seeing the object, but of seeing a presentation or an image of the object. For in the ordinary case it would seem absurd to say: "I see a presentation of the object" rather than "I see the object". But when we consider double vision, what are we to say? I see two bits of paper, but there is only one. So, taking 'see' in one way, I after all see only one bit of paper. But I can attend, point to, mean, two distinct things, A and B. Both are 'veridical'. Both are a bit of paper that I see. But they are not identical with one another. How then can each be identical with the one bit of paper that is there, and which in each case I see?

The sense-datum theory has a ready answer. I see two sense-data, each of

which is a sense-datum of that one bit of paper. Now if I stop squinting, no longer see double, do I suddenly stop seeing a sense-datum? Surely not: and what I see with one eye can't be so much altered by the fact that I am no longer seeing with both eyes. And so in any case and always I see an image, a datum. The relation must be considered between seeing the datum and *seeing the bit of paper*.

Again, consider the similar case of singular vision where there are two identical objects. This effect can be got by a simple apparatus with two eye pieces and a lens in each. A pair of similar matchboxes suitably placed appears to people with binocular vision as one matchbox. The effect is not as such stereoscopic. Parallel to the question for double vision, "Which image is veridical?", one can ask here, "Which matchbox am I seeing?" and, just as we should say that relevantly to the question *both* images are veridical, so here we ought to say that we see *both* matchboxes.

The problem about identity arises in both cases. To put it paradoxically, in the first case I see A and B, and each is identical with C but they are not identical with each other; in the second I see A, which is identical with both B and C, but they are not identical with each other. This is exactly the sort of situation that leads us to formulate new concepts for the purpose of making distinctions whereby we may avoid contradiction, or at least the paradox of making identity non-transitive. In double vision, A and B are different *images*, each of which is a presentation to sight of one and the same object C; in singular vision A is *an* image, which is a presentation to sight of both B and C, which are different objects.

In both cases there is – directly – seeing the object or objects. The conclusion is clear: it is by seeing a direct optical presentation of an O that one (directly) sees an O. An indirect optical presentation would be, for example, a reflection in a mirror or an image on a television screen. The sense-datum theorist would identify the image, which I call *an* optical presentation, as a 'datum' and would say that I directly see that, and only indirectly see the object of which it is a datum. Professor Gregory's theory would suggest that the two optical presentations in one field are a (single) perceptual hypothesis, as is also the singular presentation. I cannot accept either view.

The optical presentation has to be called perception rather than sensory input or mere sensation because it is, for example, of a matchbox and so includes 'more than is given in sense'. If we were talking simply about seeing matchboxes in an ordinary situation, Gregory would point to the matchboxishness of the perception as an indication that the perception was an hypothesis. Now when there is singular vision of *two* matchboxes, the singularity must also be hypothesis. But then also the singularity must be hypothesis when we have singular binocular vision of one matchbox. Once again, what is *not* hypothesis? The answer, it seems, now *has* to be 'input'. For even the singleness of a patch will be perception, hypothesis; and so whatever we sensorily describe will be, not sensation or the sensibly given, but perceptual hypothesis. But input is not the same thing as evidence to

which hypothesis is answerable. The lack of this is an objection to which I return again and again. *What* is framing the hypothesis? Is the conception which Gregory suggests that the framer is some mechanism that produces hypotheses which are answerable to input? For *we* who perceive don't know what the input is.

Hypotheses are *that* things are the case. Since we can perceive contrary to knowledge and belief, it looks as if Gregory's theory involved our perceiving apparatus itself as entertaining judgements. Hypotheses are predictive by way of inference. They also involve the logical constants, can be negative, universal, particular, conjunctive and disjunctive. Perception can be all of these except for being disjunctive. That is, I can see that there is *no* one in a room, that *all* the marbles in a box are red, that *some* are red *and* some are green. If I see that some are red and some green and none any other colour, that is the same as seeing that all are either red or green; but without a quantifier ('some' or 'all') perception cannot be disjunctive or alternative. By perception's being alternative I would mean two perceptions presenting themselves as alternatives at the same time. In *Eye and Brain*,[3] Gregory writes of the famous prong which starts out as two and ends as three without dividing: "The brain cannot make up its mind!" The representation is impossible, paradoxical. But there is no difficulty about an *hypothesis* being alternative. 'Or' in a report of perception not containing a quantifier is only an expression of uncertainty as to the hypothesis *suggested* by the perception.

Gregory does not distinguish between hypotheses and models. There seems to be a great difference. Hypotheses involve the logical constants, models do not. A theory that perceptions involved and were conditioned by models and were sometimes expressed by, and also suggested, hypotheses, would seem to me more promising than a theory that they *are* hypotheses.

Yet I suspect that all along I have been looking at hypotheses in a different way from Professor Gregory. Take the example of the radio in the dark room. There I could agree it might be difficult to distinguish between the perception of the radio – taking the parallel gleams as the metal trim – and the hypothesis, "That's the radio!" Or, to take another example,[4] if I am carrying dark bottles with liquid in them, I may perceive a motion of the liquid in the bottles with my hands. The statement, 'The liquid is moving,' can then be called an hypothesis, but it also expresses a perception; and here I do not know what the sensory cues are. Now as Professory Gregory has spoken of hypotheses, they sound more like schematic sketches of possibilities. Take the example of picking out language from a clutter of background noise: we might compare this with clapping a sheet with *both* some drawing *and* a complicated set of holes on to another which exhibited a great scrawl of lines. With the stencil on top, what shows through, in combination with the ready drawn lines, forms a clear picture. If I understand Professor Gregory's theory, picking out language involves some equivalent

[3] London, 1966.
[4] My attention was drawn to this by Dr A. Muller.

of having a limited – but surely very large – set of such patterns ready to clap on to presented scrawled sheets. He calls the patterns indifferently hypotheses or models, though I think he has a preference for calling them hypotheses. "Model" seems to me more appropriate.

Suppose that something like this is true for picking out language from a clutter of background noise. What else might it be true of? Of recognizing some lines in a puzzle picture as forming a face, perhaps; and, very importantly, of picking out objects as wholes distinct from one another. Perhaps also of seeing what people's movements amount to. For example, one might say, "Two people started to fight at a party I was at. Some others were holding back the one who was trying to commit an assault, but there was an element of pretence about it: he was like the Irishman saying 'Hould me back! Hould me back!' – he was really relying on their not letting him get at the other man, while thrashing about in an appearance of straining to get at him." A lot of the movement in such a scene would be irrelevant to this perception of what was going on.

For seeing the duck-rabbit as one or the other figure, or for the double or singular version which I have discussed, such an account seems quite inappropriate. It is appropriate precisely where there is something to pick out from a background in which it might well be lost. Professor Gregory has got Wittgenstein wrong about the 'tangle of lines'. That is no description of the duck-rabbit drawing as it stands by itself. The point was that the drawing might occur within a puzzle picture, where it is hidden in a tangle of lines.

7 On Sensations of Position

Wherever I spoke of 'knowledge without observation' in *Intention*,[1] I should also speak of 'knowledge without clues'. By introducing 'clues' Mr Braybrooke[2] has maintained over again what I rejected, while verbally accepting that rejection. Thus I do deny what he thinks is "a plain fact of common experience". He thinks that feelings of resistance, weight and pressure serve as clues by which one judges that one's legs are crossed, when they are crossed and one knows it in the ordinary way. Suppose they were crossed in a different position so as not to be in contact (just the calves crossed), what sensations would be his 'clues' then? Without experiment, most people would not know what sensations to expect when their legs are in that position; one might find out by putting one's legs into the position and noting what sensations one had then. This is enough to prove that those sensations are not clues by which one tells one's position.

Returning to Mr Braybrooke's case, however: if one ever did have to use the feelings of resistance in the upper leg and of weight and pressure in the lower leg as clues going by which one judged that one's legs were crossed, one would also need assurance that the sensations of pressure, weight and resistance were produced in one leg by the other and not by some quite different bodies. *Ex hypothesi*, knowledge of the position of one's legs would not itself supply that assurance.

Mr Braybrooke asks (in a footnote) what I mean by "because" when I say: "It may be because one has sensations that one knows (in the ordinary case of knowing such things) that one's leg is bent." I mean that it could be that if a leg is anaesthetized it invariably happens that the subject cannot tell whether his leg is bent without looking or feeling with his hand, etc. But compare the anaesthetic boy cited in James' *Principles*, II, pp. 489–90.

Mr Braybrooke has been misled by my consideration of the expression: "that sensation which one has on reflex kicking, when one's knee is tapped". I did not want to object to this use of the word "sensation", but argued that such a 'sensation' could not be adduced in defence of the thesis that we do after all know our bodily movements and positions by observation, because the sensation was not separable; elsewhere I implied that a sensation needed to be 'separately describable' if one observed a fact by means of the sensation. Mr Braybrooke has supposed that I was introducing a class of 'clues' which consisted in 'non-separable sensations', by means of which one tells

[1] Oxford, 1957.

[2] D. Braybrooke and others, 'Some Questions for Miss Anscombe about Intention', *Analysis* 22, 3 (1962).

From *Analysis*, 22, 3 (1962).

certain things about oneself: one's knowledge of those things, though by clues, would be without observation, because it is logically necessary that a clue by which a fact is *observed* should be sharable. Whatever the merits of this view, it is not mine; but I welcome the chance of going into this difficult topic a little further.

When I say: "the sensation (e.g. of giving a reflex kick) is not separable" I mean that the internal description of the 'sensation' – the description of the sensation-content – is the very same as the description of the fact known; when that is so, I should deny that we can speak of observing *that* fact by means of the alleged sensation.

If we are considering an expression of the form "sensation of X", we need to ask whether the words "of X" are a description of the sensation content, or whether the sensation has some different content and X is what produces or always goes with it, etc. The sensation of going down in a lift is a sensation of sudden lightness and as it were of one's stomach's lurching upwards; "of going down in a lift" is not an internal description of the sensation. It is quite possible, as is suggested by a passage of Mr Braybrooke's, that we should not have any very specific internal description of a given sensation: in "the smell of onions" for example, "of onions" is an external description, but English contains no word for the content. If onions ceased to have that smell, but it were still to be smelt elsewhere, "of onions" might become an internal description, like "bitter". People sometimes speak of a 'sensation of flying', but it is not (not at all!) a sensation occasioned by flying, and it is very unlikely that it is one that would actually go with flying: "It's a sensation of flying" or "I feel as if I were flying" is just an internal description. (The special thing about *this* kind of internal description is that it uses a word taken from elsewhere; it is as it were a metaphor – only *that* this metaphor strikes one is part of the experience it expresses. 'Sensations' of position are quite unlike this.)

The idea that it is by sensation that I judge my bodily position is usually the idea that it is by other sensations, not just the 'sensation' of sitting cross-legged, say, that I judge that I am sitting cross-legged, i.e. by a pressure here, a tension there, a tingle in this other place; such sensations are supposed to be sensations of being in that bodily position because, perhaps, they have been found to go with that. Mr Braybrooke's example has a certain plausibility for the following reason: it is natural to think of a sensation of mutual contact of sensitive parts of the body; now given the ordinary organization of the body, the contacts in question would actually imply that one's legs were crossed. If one supposes no contact, the sensations which one has will not even suggest, let alone imply, a position; unless once more one simply says: "the sensation I mean is the sensation of having one's arms stretched out in front of one", for example, where the "of" phrase gives an internal description.

In fact the sensation of mutual contact of such and such parts of the body is of the very kind that has so misled Mr Braybrooke as to my meaning; for it

is itself a sensation of position and deserves the scare quotes with which I have sometimes surrounded the word 'sensation' in this use. The reason for the scare quotes is to call in question the sense in which these 'sensations' are sensations at all. Of course we freely say: "You know what it feels like to give a reflex kick" or "to have one's legs crossed" or even "to have no sensation in one's cheek"; equally we might speak of "the sensation of" each of these things, instead of saying "what it feels like". But this only means what "the thing is familiar to you" would mean in the same context. If I imagine something when asked if I know the sensation of giving a reflex kick, I imagine giving a kick, not a sensation. If I want to imagine a sensation or feeling that I get when I give a reflex kick, and don't simply mean imagining what it is like to give a reflex kick, then I may find that I don't remember, or never noticed, what the feeling is or even where (in the knee? in the calf?). I investigate, and find that when I give a reflex kick and attend to my sensations, I get a sensation starting behind the knee and running a little way down the back of the leg, like the sensation produced by an uncomfortable degree of electrical stimulation of a muscle. That is a result of observation which I did not know before, though I should have freely said I was familiar with the sensation of giving a reflex kick.

However, not everything that is familiar readily leads us to form a phrase with "sensation of": "the sensation of being told a fairy tale" or "of building a summerhouse" sound, the one a bit extravagant, and the other wholly absurd. The descriptions which form this sort of description of 'sensation' should be descriptions of unitary occurrences in one's restrictedly bodily history. A strong reason why it is natural to call them "sensations" is that, though this is enormously rare, it can happen that one wrongly thinks one's leg is stretched out, or one has given a reflex kick or the like. Then one would say "it felt just as if I . . ." or "I had the sensation of . . ." This helps one to think that there is a sensation (corresponding to the visual impression of a blue expanse, say, having which when one looks at the sky one judges that the sky is blue) which is the datum in judging one's position, and which on occasion occurs without the position. But "I am sitting cross-legged", "I gave a reflex kick" are not thought to be couched in the language of sense-impressions, nor to be descriptions of sense-contents properly speaking, like "blue patch", "pressure", "rustle", "tingle", "pain"; and so it is supposed that the sensations of giving a reflex kick, etc., must be in principle describable in other terms which give their content and are only generally associated with the bodily positions and movements.

"I thought my leg was bent", "To me it was as if my leg was bent", "I would have said my leg was bent", "I had the feeling of my leg's being bent": in a certain context all these come to the same thing. If only my leg had been bent, there would very likely just have been that fact and my knowledge of it, i.e. my capacity to describe my position straight off; no question of any appearance of the position to me, of any sensations which give me the position. The difference between the two situations may lie only in this, that in

the one case my leg is bent and I know it, and in the other it is not bent but – off my own bat – I believe that it is. I may not even think about it at all. "I believed . . ." does not imply "I thought of . . ." or "I had the idea before my mind . . .", even when "I believed . . ." and "I had the sensation . . ." are equivalent. But if one looks for the sensation in question, what one finds is particular sensations which maybe one did not know one had, but which, under the influence of a preconceived picture of how these things must be, one now takes as the 'data'.

8 Intention

What distinguishes actions which are intentional from those which are not? The answer that suggests itself is that they are the actions to which a certain sense of the question 'Why?' is given application; the sense is defined as that in which the answer, *if positive*, gives a reason for acting. But this hardly get us any further, because the questions "What is the relevant sense of the question 'Why?'?" and "What is meant by 'reason for acting'?" are one and the same.

To see the difficulties here, consider the question "Why did you knock the cup off the table?" answered by "I thought I saw a face at the window and it made me jump." Now we cannot say that since the answer mentions something previous to the action, this will be a cause as opposed to a reason; for if you ask "Why did you kill him?" the answer "he killed my father" is surely a reason rather than a cause, but what it mentions is previous to the action. It is true that we don't ordinarily think of a case like giving a sudden start when we speak of a *reason for acting*. "Giving a sudden start", someone might say, "is not *acting* in the sense suggested by the expression 'reason for acting'." Hence, though indeed we readily say, e.g. "What was the reason for your starting so violently?" this is totally unlike "What is your reason for excluding so-and-so from your will?" or "What is your reason for sending for a taxi?" But what *is* the difference? Why is giving a start or gasp not an 'action', while sending for a taxi or crossing the road is one? The answer cannot be "Because an answer to the question 'why?' may give a reason in the latter cases," for the answer may 'give a reason' in the former cases too; and we cannot say "Ah, but not a *reason for acting*"; we should be going round in circles. We need to find the difference between the two kinds of 'reason' without talking about 'acting'; and if we do, perhaps we shall discover what is meant by 'acting' when it is said with this special emphasis.

It will hardly be enlightening to say "in the case of the sudden start the 'reason' is a *cause*"; the topic of causality is in a state of too great confusion; all we know is that this is one of the places where we do use the word "cause". But we also know that this is rather a strange case of causality; the subject is able to give a cause of a thought or feeling or bodily movement in the same kind of way as he is able to state the place of his pain or the position of his limbs. Such statements are not based on observation.

Nor can we say: "Well, the 'reason' for a movement is a cause, and not a reason in the sense of 'reason for acting', when the movement is involuntary; it is a reason as opposed to a cause, when the movement is voluntary and intentional." This is partly because in any case the object of the whole enquiry

From *Proceedings of the Aristotelian Society* (1957).

is really to delineate such concepts as the voluntary and the intentional, and partly because one can also give a 'reason' which is only a 'cause' for what is voluntary and intentional. For example, "Why are you walking up and down like that?" – "It's that military band; it excites me." Or "What made you sign the document at last?" – "The thought: 'It is my duty' kept hammering away in my mind until I said to myself 'I can do no other,' and so signed."

Now we can see that the cases where this difficulty arises are just those where the cause itself, *qua* cause, (or perhaps one should rather say the causation itself) is in the class of things known without observation.

I will call the type of cause in question a '*mental cause*'. Mental causes are possible, not only for actions ("The martial music excites me, that is why I walk up and down") but also for feelings and even thoughts. In considering actions, it is important to distinguish between mental causes and motives; in considering feelings, such as fear or anger, it is important to distinguish between mental causes and objects of feeling. To see this, consider the following cases:

A child saw a bit of red stuff on a turn in a stairway and asked what it was. He thought his nurse told him it was a bit of Satan and felt dreadful fear of it. (No doubt she said it was a bit of satin.) What he was frightened of was the bit of stuff; the cause of his fright was his nurse's remark. The object of fear may be the cause of fear, but, as Wittgenstein remarks (*Philosophical Investigations*, §476), is not *as such* the cause of fear. (A hideous face appearing at the window would of course be both cause and object, and hence the two are easily confused.) Or again, you may be angry *at* someone's action, when what makes you angry is some reminder of it, or someone's telling you of it.

This sort of cause of a feeling or reaction may be reported by the person himself, as well as recognized by someone else, even when it is not the same as the object. Note that this sort of causality or sense of 'causality' is so far from accommodating itself to Hume's explanations that people who believe that Hume pretty well dealt with the topic of causality would entirely leave it out of their calculations; if their attention were drawn to it they might insist that the word "cause" was inappropriate or was quite equivocal. Or conceivably they might try to give a Humeian account of the matter as far as concerned the outside observer's recognition of the cause; but hardly for the patient's.

Now one might think that when the question "Why?" is answered by giving the intention with which a person acts – a case of which I will here simply characterize by saying that it mentions something future – this is also a case of a mental cause. For couldn't it be recast in the form: "Because I wanted . . ." or "Out of a desire that . . ."? If a feeling of desire for an apple affects me and I get up and go to a cupboard where I think there are some, I might answer the question what led to this action by mentioning the desire as having made me . . . etc. But it is not in all cases that "I did so-and-so in order to . . ." can be backed up by "I *felt* a desire that . . .". I may, e.g. simply hear a knock on the door and go downstairs to open it without experiencing any such desire. Or suppose I feel an upsurge of spite against someone and destroy a message he has received so that he shall miss an appointment. If I describe this by saying "I wanted to make him miss that appointment," this does not necessarily mean that I had the thought "If I do this, he will . . ." and that it affected me with a desire of bringing that about which led up to my action. This may have happened, but need not. It could be that all that happened was this: I read the message, had the thought "That unspeakable man!" with feelings of hatred, tore the message up, and laughed. Then if the question "Why did you do that?" is put by someone who makes it clear that he wants me to mention the mental causes – i.e. what went on in my mind and issued in the action – I should perhaps give this account; but normally the reply would be no such thing. That particular enquiry is not very often made. Nor do I wish to say that it always has an answer in cases where it can be made. One might shrug or say "I don't know that there was any definite history of the kind you mean," or "It merely occurred to me . . .".

A 'mental cause', of course, need not be a mental event, i.e. a thought or feeling or image; it might be a knock on the door. But if it is not a mental event, it must be something perceived by the person affected – e.g. the knock on the door must be heard – so if in this sense anyone wishes to say it is always a mental event, I have no objection. A mental cause is what someone would describe if he were asked the specific question: what produced this action or thought or feeling on your part? *I.e.*, what did you see or hear or feel, or what ideas or images cropped up in your mind, and led up to it? I have isolated this notion of a mental cause because there *is* such a thing as this question with this sort of answer, and because I want to distinguish it from the ordinary senses of 'motive' and 'intention', rather than because it is in itself of very great importance; for I believe that it is of very little. But it is important to have a clear idea of it, partly because *a* very natural conception of 'motive' is that it is what *moves* (the very word suggests that) – glossed as "what *causes*" a man's actions, etc. And 'what causes' them is perhaps then thought of as an event that brings the effect about – though *how* – i.e. whether it should be thought of as a kind of pushing in another medium, or in some other way – is of course completely obscure.

In philosophy a distinction has sometimes been drawn between 'motives' and 'intentions in acting' as referring to quite different things. A man's intention is *what* he aims at or chooses; his motive is what determines the aim or choice; and I suppose that "determines" must here be another word for 'causes'.

Popularly, "motive" and "intention" are not treated as so distinct in meaning. For example, we hear of 'the motive of gain'; some philosophers have wanted to say that such an expression must be elliptical; gain must be the *intention*, and *desire of gain* the motive. Asked for a motive, a man might say "I wanted to . . ." which would please such philosophers; or "I did it in order to . . ." which would not; and yet the meaning of the two phrases is

here identical. When a man's motives are called good, this may be in no way distinct from calling his intentions good – e.g. "he only wanted to make peace among his relations".

Nevertheless there is even popularly a distinction between the meaning of "motive" and the meaning of "intention". For example, if a man kills someone, he may be said to have done it out of love and pity, or to have done it out of hatred; these might indeed be cast in the forms "to release him from this awful suffering", or "to get rid of the swine"; but though these are forms of expression suggesting objectives, they are perhaps expressive of the spirit in which the man killed rather than descriptive of the end to which the killing was a means – a future state of affairs to be produced by the killing. And this shows us part of the distinction that there is between the popular senses of motive and intention. We should say: popularly, "motive for an action" has a rather wider and more diverse application than "intention with which the action was done".

When a man says what his motive was, speaking popularly, and in a sense in which "motive" is not interchangeable with "intention", he is not giving a 'mental cause' in the sense that I have given to that phrase. The fact that the mental causes were such-and-such may indeed help to make his claim intelligible. And further, though he may say that his motive was this or that one straight off and without lying – i.e. without saying what he knows or even half knows to be untrue – yet a consideration of various things, which may include the mental causes, might possibly lead both him and other people to judge that his declaration of his own motive was false. But it appears to me that the mental causes are seldom more than a very trivial item among the things that it would be reasonable to consider. As for the importance of considering the motives of an action, as opposed to considering the intention, I am very glad not to be writing either ethics or literary criticism, to which this question belongs.

Motives may explain actions to us; but that is not to say that they 'determine', in the sense of causing, actions. We do say: "His love of truth caused him to . . ." and similar things, and no doubt such expressions help us to think that a motive must be what produces or brings about a choice. But this means rather "He did this in that he loved the truth"; it interprets his action.

Someone who sees the confusions involved in radically distinguishing between motives and intentions and in defining motives, so distinct, as the determinants of choice, may easily be inclined to deny both that there is any such thing as mental causality, and that "motive" means anything but intention. But both of these inclinations are mistaken. We shall create confusion if we do not notice (1) that phenomena deserving the name of mental causality exist, for we can make the question "Why?" into a request for the sort of answer that I considered under that head; (2) that mental causality is not restricted to choices or voluntary or intentional actions but is of wider application; it is restricted to the wider field of things the agent knows about *not* as an observer, so that it includes some involuntary actions; (3) that motives are not mental causes; and (4) that there is application for "motive" other than the applications of "the intention with which a man acts".

Revenge and gratitude are motives; if I kill a man as an act of revenge I may say I do it in order to be revenged, or that revenge is my object; but revenge is not some further thing obtained by killing him, it is rather that killing him is revenge. Asked why I killed him, I reply "Because he killed my brother". We might compare this answer, which describes a concrete past event, to the answer describing a concrete future state of affairs which we sometimes get in statements of objectives. It is the same with gratitude, and remorse, and pity for something specific. These motives differ from, say, love or curiosity or despair in just this way: something that *has happened* (or is at present happening) is given as the ground of an action or abstention that is good or bad for the person (it may be oneself, as with remorse) at whom it is aimed. And if we wanted to explain, e.g., revenge, we should say it was harming someone because he had done one some harm; we should not need to add some description of the feelings prompting the action or of the thoughts that had gone with it. Whereas saying that someone does something out of, say, friendship cannot be explained in any such way. I will call revenge and gratitude and remorse and pity backward-looking motives, and contrast them with motive-in-general.

Motive-in-general is a very difficult topic which I do not want to discuss at any length. Consider the statement that one motive for my signing a petition was admiration for its promoter, X. Asked "Why did you sign it?" I might well say "Well, for one thing, X, who is promoting it, did . . ." and describe what he did in an admiring way. I might add "Of course, I know that is not a ground for signing it, but I am sure it was one of the things that most influenced me" – which need *not* mean: "I thought explicitly of this before signing." I say "Consider this" really with a view to saying "let us not consider it here". It is too complicated. The account of motive popularized by Professor Ryle does not appear satisfactory. He recommends construing "he boasted from vanity" as saying "he boasted . . . and his doing so satisfies the law-like proposition that whenever he finds a chance of securing the admiration and envy of others, he does whatever he thinks will produce this admiration and envy."[1] This passage is rather curious and roundabout in its way of putting what it seems to say, but I can't understand it unless it implies that a man could not be said to have boasted from vanity unless he always behaved vainly, or at least very often did so. But this does not seem to be true.

To give a motive (of the sort I have labelled "motive-in-general", as opposed to backward-looking motives and intentions) is to say something like "See the action in this light." To explain one's own actions by an account indicating a motive is to put them in a certain light. This sort of explanation is often elicited by the question "Why?" The question whether the light in which one so puts one's action is a true light is a notoriously difficult one.

[1] *The Concept of Mind* (London, 1949), p. 89.

The motives admiration, curiosity, spite, friendship, fear, love of truth, despair and a host of others are either of this extremely complicated kind, or are forward-looking or mixed. I call a motive forward-looking if it is an intention. For example, to say that someone did something for fear of . . . often comes to the same as saying he did so lest . . . or in order that . . . should not happen.

Leaving, then, the topic of motive-in-general or 'interpretative' motive, let us return to backward-looking motives. Why is it that in revenge and gratitude, pity and remorse, the past event (or present situation) is a reason for acting, not just a mental cause?

Now the most striking thing about these four is the way in which good and evil are involved in them. If I am grateful to someone, that is because he has done me some good, or at least I think he has, and i cannot show gratitude by something that I intend to harm him. In remorse, I hate some good things for myself; I could not express remorse *by* getting myself plenty of enjoyments, or *for* something that I did not find bad. If I do something out of revenge which is in fact advantageous rather than harmful to my enemy, my action, in its description of being advantageous to him, is involuntary.

These facts are the clue to our present problem. If an action has to be thought of by the agent as doing good or harm of some sort, and the thing in the past as good or bad, in order for the thing in the past to be the reason for the action, then this reason shows not a mental cause but a motive. This will come out in the agent's elaborations on his answer to the question "Why?"

It might seem that this is not the most important point, but that the important point is that a *proposed* action can be questioned and the answer be a mention of something past. "I am going to kill him." – "Why?" – "He killed my father." But do we yet know what a proposal to act is; other than a prediction which the predictor justifies, if he does justify it, by mentioning a reason for acting? And the meaning of the expression "reason for acting" is precisely what we are at present trying to elucidate. Might one not predict mental causes and their effects? Or even their effects after the causes have occurred? E.g. "This is going to make me angry." Here it may be worthwhile to remark that it is a mistake to think one cannot choose whether to act from a motive. Plato saying to a slave "I should beat you if I were not angry" would be a case. Or a man might have a policy of never making remarks about a certain person because he could not speak about that man unenviously – or unadmiringly.

We have now distinguished between a backward-looking motive and a mental cause, and found that here at any rate what the agent reports in answer to the question "Why?" is a reason-for-acting if, in treating it as a reason, he conceives it as something good or bad, and his own action as doing good or harm. If you could, e.g., show that either the action for which he has revenged himself, or that in which he has revenged himself, was quite harmless or beneficial, he ceases to offer a reason, except prefaced by "I thought". If it is a proposed revenge he either gives it up or changes his

reasons. No such discovery would affect an assertion of mental causality. Whether in general good and harm play an essential part in the concept of intention is something it still remains to find out. So far good and harm have only been introduced as making a clear difference between a backward-looking motive and a mental cause. When the question "Why?" about a present action is answered by description of a future state of affairs, this is already distinguished from a mental cause just by being future. Here there does not so far seem to be any need to characterize intention as being essentially of good or of harm.

Now, however, let us consider this case:

Why did you do it?
Because he told me to.

Is this a cause or a reason? It appears to depend very much on what the action was or what the circumstances were. And we should often refuse to make any distinction at all between something's being a reason and its being a cause of the kind in question; for that was explained as what one is after if one asks the agent what led up to and issued in an action, but being given a reason and accepting it might be such a thing. And how would one distinguish between cause and reason in such a case as having hung one's hat on a peg because one's host said "Hang up your hat on that peg"? Nor, I think, would it be correct to say that this is a reason and not a mental cause because of the understanding of the words that went into obeying the suggestion. Here one would be attempting a contrast between this case and, say, turning round at hearing someone say Boo! But this case would not in fact be decisively on one side or the other; forced to say whether the noise was a reason or a cause, one would probably decide by how sudden one's reaction was. Further, there is no question of understanding a sentence in the following case: "Why did you waggle your two fore-fingers by your temples?" – "Because *he* was doing it"; but this is not particularly different from hanging one's hat up because one's host said "Hang your hat up". Roughly speaking, if one were forced to go on with the distinction, the more the action is described as a mere response, the more inclined one would be to the word "cause"; while the more it is described as a response to something as having a *significance* that is dwelt on by the agent, or as a response surrounded with thoughts and questions, the more inclined one would be to use the word "reason". But in very many cases the distinction would have no point.

This, however, does not mean that it never has a point. The cases on which we first grounded the distinction might be called "full-blown": that is to say, the case of, e.g., revenge on the one hand, and of the thing that made me jump and knock a cup off a table on the other. Roughly speaking, it establishes something as a reason to object to it, not as when one says "Noises should not make you jump like that: hadn't you better see a doctor?" but in such a way as to link it up with motives and intentions. "You did it because he told you to? But why do what he says?" Answers like "he has done a lot for

me"; "he is my father"; "it would have been the worse for me if I hadn't" give the original answer a place among reasons. Thus the full-blown cases are the right ones to consider in order to see the distinction between reason and cause. But it is worth noticing that what is so commonly said, that reason and cause are everywhere sharply distinct notions, is not true.

9 Pretending

Offered 'pretending' as a philosophical topic, I should want to distinguish between mock performances and real pretences. The difference, so far as I have noticed, is not pointed to by any of those differences between the grammatical constructions variously appropriate, sometimes to one nuance of sense and another, sometimes to one word and another closely related one, which are Professor Austin's special interest. Hence he disregards it, and lumping dissimilar things together, finds that in "the basic case" the one who is pretending must be giving a "current personal performance" in someone's presence in order to disguise what he is really doing.[1] Mock performances, to specimens of which he devotes a good deal of space, *are* most naturally exemplified in 'current personal performances' in the presence of others. But it is not at all characteristic of them to serve the purpose of disguising what the performer is really doing. That is a noteworthy characteristic of some real pretences. But for real pretences there is nothing specially basic about a 'current personal performance' in the presence of others. One can pretend to be angry in a letter (this might be mock anger or a real pretence); pretend to marry someone, the 'marriage' being by proxy; pretend to be a meat-eater in a community where vegetarianism is criminally heterodox, by having conspicuous deliveries of butcher's meat made to one's house; pretend through one's emissaries to come to an understanding with a foreign power. Whether the pretending has to be a personal performance sometimes, though not always, depends on whether the doing that is pretended has to be one. It demands a justification, which Professor Austin has not offered, to treat mock performances on the one hand, and cases like these on the other, as deviations from a centre, as fringe cases in which some of the features of 'the basic case' have disappeared. He has perhaps formed this conception out of a prejudice that the identity of a term must have something which is 'the basic case' corresponding to it.

I can at present see little intrinsic interest in mock performances. Professor Austin tells us that part of the interest of his considerations is that "philosophers who are fond of invoking pretending have exaggerated its scope and distorted its meaning". In *The Concept of Mind*[2] Professor Ryle discusses pretending, in the sense of giving a mock performance, when he prepares the ground for his attempt to explicate imagination as incipient or inhibited performance. That is a very strange account of imagination. I think it derives from the following suggestion of Wittgenstein's: suppose there were some people apparently playing tennis, but without any ball. Wittgen-

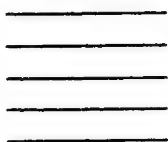
[1] J. L. Austin. 'Pretending' *Proceedings of the Aristotelian Society*, supplementary volume 32 (1958). [2] London, 1949.

From *Proceedings of the Aristotelian Society*, supplementary volume 32 (1958).

stein compared the mental image, or the calculation in the head, to this non-existent ball. We should notice that this is not the same thing as comparing *imagining* to the *mock performance* of playing tennis which is here envisaged. It is only the image which is being compared to the ball that there isn't in this game. (What would correspond to the players' strokes to and fro would be, e.g., the overt setting of a sum and the overt production of the answer.) I will not pretend to estimate the value of this suggestion, and only mention it to throw light on one of the ways in which 'pretending' has come into current philosophical literature. Obviously pretending is really quite irrelevant here. For though the tennis game without the ball could be called a mock game of tennis, and in that sense the players – in this highly fictitious example – could be said to be pretending to play tennis, the point of the example is not that this is a mock performance or any kind of pretence, but just that it is a tennis-game without a ball. And in Ryle's own attempt to describe imagination, what is of importance is the absence of something that is there when something is done, but not when it is done in imagination; it is not that such an absence is *first* supposed to throw light on pretending, and this concept in its turn is *then* supposed to throw light on imagination, as if imagination were a species that fell under it.

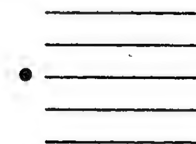
Leaving mock performances aside, let us consider how 'really pretending' comes into current philosophical discussion. Professor Austin quotes an example, about pretending to be angry.

It is fairly easy to see that the connection between the meanings of words like "pain" and "anger" and certain types of behaviour cannot be merely contingent. Just what the connection is, however, is difficult to describe in some cases. For example, it is certainly not that "He is angry" *means* "He behaves thus or thus". And yet acting a piece of typical angry behaviour might serve well as an ostensive definition of "anger". Here the inclination arises to think that if it does so serve, it is working as an indirect indication of something which is simple and yet cannot be indicated directly. This inclination arises because we remember about pretending. Let the following stand for the sort of behaviour that expresses anger:



A man may behave so and not be angry because he is pretending, and the person who understands the ostensive definition ought to understand this. Mr Bedford, in the passage Professor Austin quotes, *may* be suggesting that the question whether the man who behaves so is pretending or really angry would necessarily be settlable if only there were 'more evidence of the same sort'. And by "more evidence of the same sort" he *may* mean "more (at least ostensibly) anger-expressing behaviour" – though if he got as far as putting it like that, he would surely *not* think so.

If, then, concentrating on 'behaviour that is (perhaps) expressive of real anger' and 'the anger that it is (perhaps) expressive of', we think about pretending, we may feel forced back on a picture like this



where the dot behind the dashes stands for the anger itself. Then the dashes without the dot stand for the behaviour without the anger. This, if there is enough of it, will be pretended anger. We have to say "if there is enough of it" because, e.g., a scowling face without anger, which looks like an angry face, may be, not a pretence of anger, but just the face someone has when he is thinking hard. But there is behaviour which certainly either is the mark of anger or is simulated anger. Pursuing our picture, a plain dot without any dashes will be anger which a man does not express at all.

So, it is argued, someone who understands the ostensive definition of anger offered in an imitation of angry behaviour, will take it as an indirect indication of the dot – which cannot be directly indicated by one person to another at all. But with this conception we are forced back to the idea of the private ostensive definition with its absurd consequences – that for all we ever could know the word might stand for a different thing for different people or for the same person at different times; that we can never make more than a probable judgement that someone else is angry; or even that we cannot really make this judgement at all; that our own claim to be angry rests on an assumption that we have correctly identified something within ourselves – but without any standard of correctness – and so on.

This, then, is one great locus of the discussion of pretending. Professor Austin proposes to examine pretending just on its own account and out of the context of such discussions. In doing so he has convinced himself that a simple contrast between 'pretence' and 'reality' is no good; that pretending has such 'essential features' as that the pretender must be present and active, and there must be something, also 'on the scene', that he is disguising; that there is such a thing as 'the essence of the situation in pretending', namely 'that my public behaviour must be being done in order to disguise some reality'.

Against this I would argue that pretending can no more have that type of 'essential feature' than falsehood or identity or seeming can. Seeming is especially relevant, because the notion of pretending is closely bound up with that of seeming. The best general account of pretending would be something like: *the production of a would-be seeming to be*[3] *what you are not*. That is clumsy, so I will shorten it to "trying to appear what you are not": cases of this which

[3] Like Professor Austin, for brevity's sake I disregard verbs other than "to be" in formulating this.

would not fall under the longer form are excluded. The point of this exclusion is that a man might try to appear what he is not, and not succeed in doing anything – e.g., a very sick man, trying to seem cheerful and too weak even to smile, would have only tried to pretend.

From this general account of pretending we can see why the two more specious implications criticized by Professor Austin do not hold. As he says, pretending does not imply not being, and really being does not imply not pretending. A man can pretend to be poisoned when, unknown to him, he is poisoned. Now, in "trying to appear what you are not" the words "what you are not" are governed by the "trying": the whole phrase does not mean: "concerning something which you in fact are not, trying to appear that thing", but: "trying to bring it about that, without being something, you appear that thing".

This general account of pretending needs an addition to include some cases of trying to make it seem that something is the case which is not. For example, one might pretend that one's child was under three years old (to avoid paying a fare) by having him dressed in rather babyish clothes and carrying him like a rather younger child, as well as by what one said. All these details would be part of the pretence. In such a case, we have to speak of 'pretending that' rather than 'pretending to' because the subject of what is pretended is not the same as the pretender, and not, I think for any other reason. Two central features of "pretend" are: (1) that the pretender should figure as a principal, in what is pretended *and* in that by which it is pretended; I mean the latter in such a sense that he would *be* a principal if the appearances were not deceptive. This condition may be satisfied even if he is not where the pretence is carried out, if what is done is something that could be done, with him as principal, without his presence, as in the case of the king coming to an agreement with a foreign power. My corollary (that he would be a principal in that *by* which the pretence was made, if the appearances were non-deceptive) can be seen to be necessary from this: if, for example, the king arranged a deceptive appearance that his emissary proposed to murder him, he would not thereby be pretending to be a proposed victim of assassination. (2) Further, there is what might be called a 'rule of sequence of tenses' for "pretending"; if someone *has* broken some crockery and left it about so that I *shall* think he *was* angry, he *was* not pretending to be angry; and, unless he does something *now* to exhibit the smashed crockery as the result of past rage on his part, he *is* not now pretending *to have been* angry. I suspect that such facts have influenced Professor Austin; he has construed them as a necessity for the pretender to be 'present on the scene' and 'giving a current personal performance'. – In the pretence of the fraudulent traveller with the child, the traveller *is* a principal and *is* acting. I will not consider such cases further; though one has to speak here of 'pretending that' and not 'pretending to', this is only because of the diversity of subjects, and such cases of 'pretending that' should be subsumed under 'pretending to'.

'Pretending' is an intention-dependent concept; one cannot pretend inadvertently. But no special further intentions in whatever constitutes pretending in a given case are specially basic "*as Professor Austin pretends*". – Why would that be rude and unfair? Because it implies that he has been trying to make-things-seem-as-they-are-not. There is no hint in that piece of rudeness that the publication of his paper serves to disguise something he is really up to, and it is not the absence of such a hint that turns it into a fringe use of "pretend".

'Seeming' can have no 'basic case'. Let A be the subject of a predicate x. Then we can ask "What is it for A to x, or to be x?" and further "What is it for A (only) to seem to x or to be x?" This latter enquiry may well throw light on the first question. And we *could* ask further "Can A be so responsible for phenomena by which he (only) *seems* to x, that it accords with the grammar of "pretending" to say he pretends to x?" In cases where that is so, an investigation of 'pretending to x' will often help us to understand the concept 'x' better. But the quite general characteristics of the verb "to pretend" are likely to give singularly little light in an enquiry into 'pretending to x'; such an enquiry must be completely dominated by the character of the 'x' in question.

In the case in hand – that of pretending to be angry – if we consider when and why we may judge that someone was only pretending, we see that it is not only features of his ostensibly angry behaviour that prompt the judgement. If it were, then 'being angry' would be much more like, for example, 'feeling jumpy' than it is. Pure pretences of being angry in person are rarely so successful that a discerning judge will not detect them in the tone and expression of the subject. However, such admirable pretences are possible; so of course the philosopher supposes a case where the performance is perfect. Then perhaps he feels driven either to such a recourse as Mr Bedford's – "there is a limit that pretence must not overstep" – or to postulating something hidden behind the behaviour. But, as Professor Austin indicates without enlarging on it, there is more to look for besides giveaways in behaviour. Anger has four main features: (1) its object, (2) its expression, (3) feelings, (4) aims. By "angry behaviour" we usually mean things falling under (2), the expression of anger: the angry-looking face and gestures, the stamping or trembling or rigidity, the tone of voice, perhaps the pointless smashing of things. (2) may include elements that bring in (1) and (4). If an angry man expresses his anger in speech, his speech will probably characterize the thing or person or situation or spectacle that he is angry with either as bad in some way or possibly as something to be overcome or resisted. I suppose that is why Aristotle said that anger was more 'rational' than lust – the expression of anger by an angry man often gives *grounds* of anger. A story of anger – real or pretended – usually includes what the anger was at or supposed to be at, so characterized that the hearer can understand it *as* an occasion of anger. For example, if a man is said to have been angry at the sight of a chair, in a way we do not yet know what he was angry at; we

need an explanation which will make clear whether his purposes or orders have been frustrated, or his vanity insulted, or someone has been proved to have behaved abominably – or what.

There are also characteristic aims of anger – to harm or afflict someone or something, or to overcome obstacles[4] or resist or repel something. A man who was careful to give no sign of anger and did not even have specific angry feelings (sensible commotions) might be implacably angry and arrange some way of harming the man who angered him. Thus, though (2) may pass into (4), as when someone immediately starts strangling the person he is angry with, there *can* be a great difference between the expression of anger and its aims. A man could be said not to have given expression to his anger at all – he merely brought it about that the man who had offended him was ruined or hanged.

What is feeling angry? Let us suppose we find someone who has just been angry and ask him what he felt while he was angry. He may well say, e.g., "I felt hot", "I felt cold and trembling", "I felt a rush of blood to the head", "I felt a slight tension in the chest". Yet feeling angry is not any of these things; otherwise we could produce the sensations he characteristically has when angry – e.g., by means of some electrical apparatus – and say: "There, now you feel angry." On the other hand, those sensations were not just concomitants of his anger; he might feel something else, a pain in the stomach, let us say, while he was angry, and not mention it as 'what he felt while he was angry'. The sensations that he mentions are the ones he – intuitively – gives as what he felt *in* being angry. Or again, we may say that he gives his anger as an interpretation of those sensations. But is there nothing else that the *felt* anger is? One kind of reply to this might be: "I felt: 'You filthy swine!' or 'This is too much!' or 'That trick again!'" The words, or the thoughts, are themselves an angry reaction, and there is no need to postulate, indeed no sense in postulating, another reaction, not the words or the thought, which is the ground of the words or the thought and is the felt anger itself. The fact that the verbal reaction may be a sham does not prove such a need. The

[4] I owe notice of this aspect of anger to Plato, made intelligible by Aquinas who adopted this part of Platonism, getting it apparently from St Gregory of Nyssa and St John of Damascus. He does not have a tripartite division of the soul like Plato, but divides the 'sensitive appetite' into two parts, the 'concupiscible' and the 'irascible'. Through the one, he says, the animal is simply inclined to pursue what it needs and to flee what is hurtful, through the other to resist what attacks its needs and offers hurt to it. "These two inclinations do not reduce to a single principle, because the animal sometimes faces hurt against the inclination of desire, so as to oppose what opposes it according to the inclination of anger. Hence the passions of the irascible are even seen to be at war with the passions of the concupiscible. For in general *as desire burns higher anger sinks, and as anger burns higher desire sinks.*" (*Summa Theologica*, Ia, 81, art. 2). To understand the force of this remark we should imagine someone, about to engage in sensual enjoyment, having to fight to retain what he wanted to enjoy; and then, the battle won, returning to engage in enjoyment. But, Plato might say, anger is not uniquely concerned with sensitive appetite. That is because of our organization: "while I was musing, a fire kindled". An abstruse thought can bring my fist crashing down on the table and so also cause all sorts of reverberations in my sensuality. Hobbes' definition of anger as "sudden courage" must be in this tradition.

mistake is to suppose that since a man can say "I felt angry" we shall find out what anger is by finding out what he felt. In what context does he say "I felt angry"? In the context of some story of events, conversations, thoughts: that is to say, he puts the anger he reports into a context which shows a lot about the anger: and what it shows is not just extraneous. That is why looking for the meaning of "anger" in what a man feels who feels angry yields such dissatisfying results, as if the anger itself had slipped between our fingers and we were left with details, which, while relevant, do not add up to anger.

I am not saying that every case of anger must have all these four features – rather, here we do have a 'full-blown' sort of case with all these features, and other cases lacking some of them. Now imagine an anthropologist saying "The psychology of this tribe is odd: they are angry only, and always, before sitting down to a really good meal." Asked why he says so, he explains "Then, they always shake their fists and assume an expression of hideous rage; after that, they sit down to eat; and they never shake their fists or assume that expression at any other time." – Would it not be absurd for someone so much as to say "They must be only *pretending* to be angry"? Once we have recalled these points about anger, we can see how a diagnosis of pretence could be made in face of angry behaviour which was a quite flawless performance. For example, one might know that the man did not really mind about what he was ostensibly angry at; that it really suited his book extremely well and that he knew this. Or that the supposed affliction that he was laying on the victims of his anger was not really an affliction at all but something agreeable and that he knew this.

If someone claims that he was only pretending to be angry on an intelligible occasion for anger and when his performance had been good if it was only pretence, it is natural to ask why he was pretending that; and an answer telling more about the situation, his attitudes and what he was after will help to convince us that he was pretending.

These facts point to one great difference between anger and pain, and generally between passions and sensations. If a person's performance is good and – as may be the case – there is nothing else to look at, there may be no way at all of telling whether his pain is sham or not, if, say, it is a brief pain or he does not make the mistake of behaving inappropriately when, e.g., he does not think he is observed. But it is absurd to say (as Mr Bedford says in the passage quoted by Professor Austin; which shocks me, though not Professor Austin) that he alone is in a position to give decisive evidence! What he says is no more decisive than his behaviour is. If one thought his groans might be shamming pain, one would hardly accept his word. This however does not mean that there is quite generally a difficulty about knowing whether someone is in pain or not. The difficulty occurs in some cases; and sometimes cannot be resolved. Cases can be constructed for anger too; but there is much more to consider in cases of anger: the whole story of the occasion ('whole story' in the sense of "whole truth" in the law-court oath). Contrast "As I walked along the passage I had a sudden stab of pain in

my chest," and "As I walked along the passage I had a sudden stab of anger." Anger what at? "Nothing at all." This man is talking nonsense – unless he means "At X, which I judge to be a nothing." On the other hand consider this case: an actor, who has to act an angry man in a play, says "When I act it, I really am angry." He backs this up by saying that he *feels* angry, and he *means* the angry words in which he recalls and threatens evils. Would not a dispute be stupid about whether he is correct to use the words "I really am angry" or not? Say which you like, so long as you are clear about the facts. *This* situation does not arise for physical pain. For if an actor in *King Lear* said "It's a most extraordinary thing, when they tear out my eyes, I feel an agonizing pain as if it were really so, I almost think I shall have to give up the part," well, we believe him or not, there is not a choice, *after* we believe him, between saying "He really feels pain" and "He doesn't really".

Although I have given reasons for accepting Professor Austin's remark that pretending does not imply not really being and really being does not imply not pretending, I have the impression that his own reasons for saying this lie at least partly in his examples, such as that of the man who was cleaning the windows and at the same time 'pretending to be cleaning them'. Here he relies on a nuance which he explains to us. (It may not exist everywhere where English is native.) But the whole reason why a man can be said to be pretending to be cleaning windows (when he also is cleaning them) is that what he is pretending is not the case. The explanation of the nuance makes this clear. The observer diagnoses the window-cleaner's felonious interest and guesses from this that the window-cleaning is a fake. The diagnosis might be right but the guess wrong – if, say, the man were the regular window-cleaner doing this regular job on his regular day. Professor Austin explains "It is still a pretence [i.e., though the windows *are* being cleaned], because what he is *really* doing is something quite different." But the point of the expression "What he is *really* doing is something different" is that 'what he is *really* doing' falsifies the appearance he presents by cleaning the windows. There are other things he might also be 'really doing' – such as earning his wages or composing verse – which would also be 'different' from window-cleaning but which don't falsify "what he is really at is cleaning the windows" at all. The appearance presented by cleaning the windows is that, in cleaning the windows, he is doing something in some ordinary and proper course of things; and that this is a false appearance is the meaning of the expression "he is pretending *to be cleaning* the windows" in this context.

The two sentences

(1) He is cleaning the windows,
(2) He is pretending to be cleaning the windows,

may both be true; and as a matter of grammar "is cleaning" is the indicative corresponding to the infinitive "to be cleaning". Does Professor Austin think that *this* is therefore a counter-example to "pretending implies not really being"? And is it perhaps a fairly important step in his argument, enabling him to reject 'false appearance' as quite central to pretending? If so, this is grammatical superstition.

Why cannot a baby six months old pretend to be in pain? A mother might say "The baby pretends", and we "You mean there's nothing wrong, it only cries to be picked up." Suppose she insists that there is more to it, the baby is a clever one and *really* pretends? Mothers and others sometimes talk nonsense of this sort. The question is how we know it is nonsense. It is not competence to perform a mental act of pretending that is in question. Wittgenstein would say "Pretending is part of a complicated form of life which the baby is not living yet," but what does that mean? English people are apt to say "The dog is pretending to be lame." Why? He limps, but if he sees a rabbit he rushes after it with no trace of a limp. He was lame and got a lot of special kindness, and is looking for more. We assimilate this behaviour to human pretending. Once these facts have been stated it is not a further hypothesis that he is pretending. The behaviour of the baby is not like enough for the assimilation to be attractive except to mothers, etc. But what is it not like enough to?

The answer to the questions raised here is that you cannot ascribe real pretence to anything unless you can ascribe to it (1) a purpose and (2) the idea 'can be got by seeming to – '. That is why the baby case is nonsense; the baby's purpose may be clear enough, but what reason could there be to ascribe to it more than the idea 'can be got by roaring'? And even this means no more than that the baby roars *to be picked up*. Then why should we say more of the dog than that he limps to be petted? Why indeed? Only because limping has such a characteristic appearance, is not just going on three legs but has an air about it, so that if the limping is voluntary, we may implicitly think of the presentation of this appearance as deliberate. We have once more reached a point where we should say "Say 'he's pretending' if you like, or refuse to if you like, so long as you are clear about the facts." I emphasize this; because I am not sure whether Professor Austin would ever admit that we ought to say "Say such-and-such if you like, so long as you are clear about the facts"; if he would have some objection to this, I should like to see it brought out into the open.

These considerations yield this result: we sometimes ascribe pretence by way of a comparison, a sympathetic projection on to a body of facts which we compare with some of the facts of fairly developed human life. Apart from such sympathetic projections we must say: we can only ascribe pretence to beings to which we can also ascribe purposive calculation. That is not because pretence is generally purposive. It is not; wanting to seem something that one is not, without any further end in view, may even form the biggest part of pretending. But it must be significant that when we ascribe pretending to animals, it is because we see an advantage gained by seeming. Without meaning anything absurd (like the mother) we find it possible to speak of animals', birds' and insects' pretending to be boughs, leaves, twigs, etc. I think this shows reason to speak of purposive pretending as 'basic'.

When we consider unpurposive pretending, a new distinction appears between what I will call plain and non-plain pretending. Unpurposive pretence may be 'just for fun' or 'to tease' and the like. The description "unpurposive" may be challenged on the ground that teasing or fun is a purpose, but I think the challenge would be wrong. It is a specific advantage served by seeming that is characteristic of the purposive pretending that is 'basic'; fun and teasing are something one diagnoses as one diagnoses dancing or playing a game, not by seeing them as results achieved in a certain way. I will call pretending "plain" when the pretender unreflectively knows that he is pretending. A great deal of unpurposive or only very vaguely and diffusely purposive pretending is non-plain.

What I have in mind is best illustrated by an example. Here is a dialogue between a schoolmaster and a parent summoned for interview:

Did James tell you I had to beat him to-day?
Yes, he said he got beaten.
Oh, did he tell you what it was for?
He told me it was for something he had written in his book.
Hm! I don't suppose he told you what he had written.
I don't know – what he said was that he wrote "Casson is a sod". I gather Mr Casson is one of the masters.
Oh! . . . Well, that's not very nice, is it?
Well, I understand your beating him, but all the same, surely this is quite an ordinary thing for a boy to do?
No, in my experience, not at all normal.
Let the parent's reply to this be unspoken, since it is: "Stop pretending".

In this example, it might be tempting to call the schoolmaster's last remark a plain lie. But we ought to notice that most likely that is just what it is not. A lie is a plain lie when it contradicts what the speaker unreflectively thinks. I do not mean "when it contradicts an explicit thought" since (as is well known) 'what a man thinks' is not the same thing as 'what he *is at the moment thinking*' – even if it is only what he thinks for the time being, in the particular context. But sometimes it would take some reflection, in the circumstances, for a man to realize that he knew the contradictory of what he said. Then what he says is not a perfectly plain lie; he can even be said to think it.

It is not, however, his saying what he knows to be untrue that makes our schoolmaster's case one of (non-plain) pretending. He could pretend in this sense without saying anything untrue. Further, we often tell untruths that are not lies, in the sense that they do not contradict what we *unreflectively* know to be true, without 'pretending' in any sense beyond 'making out true what (we know) is not'; and, where the content of the 'pretence' is just the content of what is falsely said, there is no particular aptness about the word "pretending". We say a thing when we know it is not true, and yet without telling plain lies, in many ways; one is, by falling into cliché. For example, a sufficiently learned author speaks in a popular book of Hobbes' "militant atheism"; we are in a position to know that he knows that if Hobbes was an atheist he was a crypto-, not a militant, atheist. But the fact that he wrote that phrase decidedly suggests that it would cost him a brief moment's recollection to realize that he knew this. Here, however, a use of the word "pretend" really would be a fringe use, as applied to a single statement; it would come to nothing but: "He says so-and-so, which he must know not to be true." But there is a sense in which the schoolmaster is pretending which goes beyond his telling a (non-plain) lie. What is in question here is *hypocrisy*: and we are trying to make out what kind of pretending this is.

The following example brings out the contrast between mock performance, plain pretence and hypocritical pretence. A certain nun was the heroine of a devotionally exciting story; the story was generally known, but not her identity. Once someone guessed and said "So you are the one!" She, 'with such simplicity' – so the story runs – 'that the other was completely deceived', laughed and said "So you have found me out!" Thus she was pretending to be making a mock admission of something – with a view to concealing that it was the case. This, then, was a plain pretence. The word "simplicity" bears dwelling on. It does not merely mean that she laughed and spoke in a natural way, just like someone who really was making a mock admission of something that was not the case. Nor can it mean that she acted without guile, for the contrary is being recorded. With this word the story-teller is insisting that the pretence just was a genuine concealment of her identity, and not *itself* a further pretence of a new sort, as it were saying "See how I am one who wishes to remain obscure". The story-teller probably wishes to suggest that the episode marked a genuine wish to remain obscure; not a pretence of having such a wish. *This* pretence, if the wanting-to-seem was just for its own sake, would be not plain but hypocritical pretence. It is characteristic of this sort of wanting-to-seem that it carries with it an implicit demand for respect for an atmosphere evoked by the pretender, which surrounds not the reality, but the *idea* of such things as being principled, or cultured, or saintly, or rich, or important. There is something of which the schoolmaster is as it were saying 'Respect this'.

This throws light on a further notion, one of the popular senses of *cynicism*. In my sense of "plain", this is a 'plain' pretence of hypocrisy, and is found, e.g., among the clearer-headed politicians.

10 On the Grammar of 'Enjoy'

What does the verb "to enjoy" take? An object, as in "He enjoyed the joke"? Sometimes; but should we speak of it as 'taking an object' in "He enjoyed compelling Jones to support the motion"? Or in "He enjoyed swimming in the lake"? Here we have complex expressions which include verbs, i.e. "compelling" and "swimming". Someone may feel inclined to say that this complexity is the sign of a proposition and that, logically speaking, we have the right to say: enjoyment is here enjoyment *that* such-and-such is the case just as belief is belief *that*. . . .

There is one objection to saying this which is really a misunderstanding. It sounds ludicrous to speak of enjoying a proposition, except in cases where that is like being amused by one. That would be the same sort of thing as enjoying a joke or a tragedy. When I, say, enjoy talking to people, or again enjoy being away from everybody, I don't enjoy a *proposition*! But why not? "Being away from everyone" – that's a description of a state: we have at least as much right to speak of a proposition being taken by the verb "to enjoy" here as of a proposition being taken by the verb "to believe" in "I believed I was away from everyone". It will seem wrong only if our conception of believing is that it is a relation to or operation on some sort of abstract entity intermediary between the subject and things. But we need not have that conception.

The idea we are concerned with is this: if, in a report of enjoyment the expression telling what was enjoyed is a complex expression containing a functioning verb, then the enjoyment must be understood as enjoyment *that*. . . . This idea is a symptom of what we might call "the *Tractatus* in the blood stream" – an affection to which no one in philosophy of a certain tradition is quite immune, whether he likes it or not. We find in the *Tractatus*:

> Wherever we have complexity, we have argument and function, and that means we already have all the logical constants . . .(5.47).

To say that we have "argument and function" is to say that we have a predicative connection of expressions. What is this, and when do we have it? The passage says: wherever we have complexity; but all that is implied is that when there is complexity, then there is a proposition *in the offing*. The conception that we are considering at the moment is that the complex expression telling what was enjoyed should *itself* be explained as, or be rephrased as, a proposition. For it typically contains a *verb*, and that gives us a predicate or function. If the verb has indeed such a role, then *here*, in the verb and whatever completes it, we must have before us at least the *equivalent* of a proposition – what amounts to one.

From *Journal of Philosophy*, 64 (1967).

In logic we assimilate other verbs to "being" with its complement, and we call them all predicates. We use a notation in which we write $F(a)$ equally for "*a* is *F*" and "*a F's*". We would even treat 'becoming *F*' in the same way, putting $B(a)$ where $B =$ "becomes *F*".

There is a deep ground for this assimilation. It is to be found in the predicative mode of connection of expressions. Consideration of the verb "to enjoy" will help to bring this out. There *is* such a thing as 'enjoyment that. . .'. The expression is not good English; but if we insert the words "the fact" or "the idea", there is no objection to it. Thus I may have enjoyed, taken pleasure in, the fact that I was riding with *N*. But that is not the same thing as actually to have enjoyed riding with *N*. The fact may have given me pleasure, but if I am candid I may have to confess that I did not enjoy the activity itself.

We may therefore use the very fact that there is such a thing as enjoyment *that* . . . against saying: where what is enjoyed is described by a complex expression containing a functioning verb, enjoyment is always enjoyment *that*. . . .

Now the contrast between enjoying *F*-ing and enjoying the fact that one is *F*-ing does not remain in being where we put a *negation* into the phrase telling what was enjoyed – I will call it the *enjoyment-phrase*. If I enjoyed *not* swimming, that is the same thing as to have enjoyed the fact that I was'nt swimming. "Not" is one among many logical signs. There are the other truth-particles, i.e. the truth-functional connectives, and there are the many quantifying applicatives, like "some", "only", "most" and "the" where this is the mark of a 'definite description'. Any of these may occur in the enjoyment phrase. When they occur, they can sometimes be displaced, i.e. be put outside the verb "to enjoy", leaving the whole sentence equivalent to the one we started with. But sometimes they can't. Thus, on natural interpretations, the quantifier "all" is displaceable in

> I enjoyed talking to all the people in the office

but not in

> I enjoyed being admired by all the people in the office.

For the former is naturally taken as equivalent to

> Of each of the people in the office it held that I enjoyed talking to him.

But, by the most natural way of taking the second report, the "all" must remain inside the enjoyment-phrase.

Similar contrasts can be found for truth-functional "if" and "and". I had a splendid holiday, I enjoyed fishing and climbing and riding and playing bridge. Perhaps this is equivalent to "I enjoyed fishing and I enjoyed climbing and I enjoyed . . .". But perhaps not. If not, we have here a case of enjoyment *that* . . . (Note, however, that in "I enjoyed lying in a hot bath and reading" the "and" is not truth-functional.) An interesting example was

given me by Sue Larsen: "I enjoyed paying by cheque if I wanted to." This is enjoyment of that fact or idea; i.e. it is enjoyment *that*.

With these considerations in view, we may adopt as the mark of predicative connection the admissibility of three things – any or all of them; it doesn't matter, for they go together. To have a verb – or other predicable expression – connected to a further expression by the predicative mode of connection, so that the verb becomes a *predicate* and the further expression is a subject-expression, means that we have the *admissibility* of

(1) the combination of two such structures by truth-connectives;
(2) the introduction of quantified expressions – e.g. their substitution for proper names;
(3) the introduction of negation.

What I mean by "admissibility" is this: these procedures don't make a difference to the *type* of thing being said, like the difference between "I enjoyed *F*-ing" and "I enjoyed the fact that I was *F*-ing". Thus, suppose I 'enjoyed talking to the most handsome person present'. This is so far ambiguous – was it just an accident that he was the most handsome person present? That is, was it just that this man *was* the most handsome person present, and I enjoyed talking to him? I don't mean: was it an accident that he was handsome? But rather: does the definite description "*the most* handsome person present" belong undisplaceably in the enjoyment-phrase? If so, then the enjoyment phrase manifests not just a *significant* connection, but a *predicative* connection of expressions, and we may say that here is a case of enjoyment *that* . . .

When we have a predicative connection of expressions, we can discern in the whole that is so connected at least one part which is a predicate and discern, or supply as implicitly present, at least one part which is a subject-expression. (An example of the subject-expression's being implicit is supplied by "I enjoyed not swimming".)

I am saying that, if the definite description belongs properly in the enjoyment-phrase, "talking to the most handsome person present" is signalized as a predicative connection of expressions by the mere fact that it contains a definite description in significant connection with the verb "talking to". There are several points to dwell on here.

First, significant connection is of course more than mere juxtaposition. No doubt we grasp significant connections of words if we use and respond to them – e.g. answer them, or obey them, or act on them, or pass on their gist. But it is more sophisticated to have the *concepts* of significant connections. One is taught these concepts by being trained in grammar.

In grammar we acquire the concepts of subjects and objects of verbs by being presented with sentences containing verbs, say "He ran", and being trained to answer the question "Who or what ran?" by repeating the word "He". Or, given the sentence "John kicked Jim" and the question "Whom or what did John kick?" we answer, "Jim".

We can reasonably extend the notion of a verb so that it includes whatever extras it is necessary to quote in asking the "who or what?" or "Whom or what?" questions. Thus, I have called "talking to" a verb, and would call "swimming in" one.

The teacher who asks such questions knows how to ask them – e.g. he doesn't ask "John who or what Jim?" And while he himself may have been taught in turn by a similar training, some original grammarian must have leapt to understanding, to the formulation of the concepts of subject and object, in framing these questions. Such a question is not necessarily apt just because we have a functioning verb. For example, given a sentence "It's not my custom to swat flies" the question "Who or what to swat flies?" ought not to be asked, and certainly not in order to get the answer "custom" or "It's not my custom". The question would be an ineptitude.

This is sometimes so even when the verb is in the indicative mood. Take for example "Anyone who eats cheese is a scoundrel". The question "Who or what is a scoundrel?" goes through smoothly; we are prompted to give "Anyone who eats cheese" as an answer. But the questions "Who or what eats cheese?" or "Who or what eats cheese is a scoundrel?" should not prompt an answer "Who" or "Anyone who".

We may compare learning to answer this sort of question to learning to answer the question "What letter comes after L in the alphabet?" Only there is more intelligence in it, for it is not a matter of merely repeating, say, the whole of what precedes the verb in a sentence or the whole of what succeeds it. Thus, given "Smith is a ninny and James is a dolt" and asked "Who or what is a dolt?" the learner answers "James", not "Smith is a ninny and James". But, given "The man next to James is a dolt" he answers not "James" but "The man next to James".

At the bottom of our understanding in innumerable cases there lies a catching on to what is done which is not dictated to us by the elements of the situation. (If this catching on were to be *explained* by an inborn capacity, we should have before us, in some form or other, a doctrine of innate ideas. Someone who wishes so to explain it has to consider, then, what he may mean by a doctrine of inborn capacity, innate ideas, For if inborn capacity is an explanation, it cannot amount to the mere fact that we do master language.)

We have an example of significant connection, then, when subject or object can aptly be asked for. But now, so long as we are dealing with predicative connections of expressions, there is no philosophical point in making any distinction between subjects and objects of verbs. We can assimilate a sentence containing subject, verb, and object to a sentence stating a relation. And just as we may call not only

was to the right of *B*

But also

A was to the right of

a predicate, occurring in

A was to the right of *B*

with "*A*" as subject-expression in the first case and "*B*" in the second, so also in

Mary was talking to the most handsome person present

we may call

Mary was talking to

a predicate, with "the most handsome person present" as a subject-expression connected with it. (The point will be familiar to students of Frege.)

Thus – though we are here not following the terminology of high school grammar – if, singling out the enjoyment-phrase "talking to the most handsome person present" I ask "talking to whom or what?" the correct answer can be said to give a subject and the connection to be predicative, *if* the enjoyment-phrase contains the definite description undisplaceably.

If, however, the definite description is displaceable, i.e. if we produce an equivalent report, as the sentence was meant, by placing the definite description outside the enjoyment-phrase, then we do need the distinct concept of an object of the verb in the enjoyment-phrase. For it may be that, though – of course – the expressions in the enjoyment-phrase do stand in significant connection, they do not stand in predicative connection: the enjoyment is not enjoyment *that* . . . Where a purely logical sign belongs essentially in the enjoyment-phrase or could be introduced without altering the kind of thing being said, we have enjoyment *that* . . . Where this is not so, we have enjoyment of activities themselves, or things, or happenings, or existence itself. I will call this "enjoyment of substance" and the other kind "enjoyment of fact".

It will be wise to remark at this point that both kinds may be illusory – the fact (as we say) not a fact, the activity not one that is indeed taking place. This is a large topic on which I don't want to embark here, but the observation is necessary lest for example a case of enjoyment *that* . . . should seem to be wrongly called "enjoyment of fact" if the that-clause was false. We do in any case need a qualification here, however: a usual one would be a parenthetic "as he thought", or "as I thought" – e.g. "What I enjoyed about the occasion was that, as I thought, there were no Left-wingers present". 'Enjoyment of fact' is a notion which I would want to cover this sort of case too.

Enjoyment of substance is much more important than enjoyment of fact; perhaps neither occurs pure among those who have language; perhaps at least enjoyment of fact is always grounded in *some* enjoyment of substance. But my immediate aims and interests in this paper are rather restricted: I am interested in the characterization of predicative connection. It is in a way an accident that the starting-point of the inquiry concerned pleasure. It is not

an accident, in the following way: the different logical characteristics of different types of enjoyment-phrase are what have caused me to distinguish between the predicative mode of connection and another, but still significant, mode of connection of words. I suspect that enjoyment of substance may be compared to perception; that the phrases "a man swimming", "a red circle", occurring, as they may, as object-phrases in reports of perception, are further examples of nonpredicative connection of expressions. Certainly there is a mysterious difference between "a man swimming" and "a man is swimming". This is not to be explained – as one might at first blush be tempted to explain it – as the difference between, say, concept and judgement, the latter being qualified by an *act of acknowledgement* (Brentano's *Anerkennung*) absent in the former. For "a man is swimming" may occur as, e.g., the antecedent of a conditional, and then it isn't 'judged' or 'acknowledged'. Nor does the "is" seem to be the essential thing that is important here – we could imagine "a man swimming" being used to make the assertion or as the antecedent of a conditional. If it were so used, nothing would seem to have got left out.

What is in question seems to be the distinction between a concept or term and a proposition. It will be remembered that our whole inquiry opened with the question whether we should think of the verb "to enjoy" as 'taking' what ought properly to be understood as a proposition. But I would not choose to call even cases of enjoyment *that* . . . cases where the verb takes a proposition. The reason is that the reflexive pronoun, which would usually be at least implicitly present as subject-expression, makes a mess of the idea of a proposition here. For I assume that a proposition should be an independent unit. But "*A* believes that he himself . . ." doesn't seem to be a substitution-instance of "*A* believes that *x* . . .". For it is not equivalent to "*A* believes that *A* . . .", since the two might have different truth-values when completed in the same way. *A* might not know that he was *A*.

Thus, only if there is such a thing as my enjoying the fact that *N* . . ., am I willing to speak of "enjoy" as taking a proposition. Here one would have to inquire whether, in order to make the thought explicit, there isn't needed an insertion involving myself or an activity or state of myself, before "the fact that *N* . . .". For example, suppose that I enjoy the fact that *N* is discomfited. Does that require a construe as:

I enjoy thinking of the fact that *N* is discomfited

where 'thinking' is an activity? or as

I enjoy knowing that *N* is discomfited

which I would argue *is* a case of enjoying *that* . . . , but a kind that involves myself, i.e.,

I enjoy [the fact] that I know *N* is discomfited.

Thus it appears problematic whether we should *ever* think of "enjoy" as 'taking a proposition', even though we are content to think of it as sometimes taking a that-clause. 'That' doesn't necessarily serve to introduce a proposition.

A standard conception of a proposition is that it is an instance of a certain sort of complex of signs (or what is expressed by that) which is such as to be true or false. Now in putting forward my suggestions about predicative connection I am, as it were, turning the whole thing round: I am saying that a predicative connection of expressions *is* just the sort of connection into which we introduce negative signs and quantifying applicatives, and which we combine with others by truth-connectives.

Part Two

Memory and the Past

11 The Reality of the Past

The first statement of the problem of the reality of the past is in Parmenides. "It is the same thing that can be thought and can be," so 'what is not and cannot be' cannot be thought. But the past is not and cannot be; therefore it cannot be thought, and it is a delusion that we have such a concept.

The doubt raised here is not a Cartesian doubt which is meant to be a doubt about facts; the question is not "May not all our apparent information about the past be incorrect?" but "Is not our apparent concept of the past impossible?" Parmenides' remarks suggest the enquiry "How is it that statements about the past have meaning?"

We are not concerned here with any other objects of thought besides particular things, events, circumstances, places, persons and so on. That is, not with abstractions, generalizations, imaginary things. When I think of my acquaintance *A*, and think that he is in Birmingham, it is he, *A*, the very man himself, and Birmingham, that very place, that I mean, and not some intermediate representation of them. I might try to emphasize this by going and finding and pointing to the man and the city; not that I imagine that I should thereby make clear what it is to mean them, but I should then be exhibiting them themselves, and I want to insist that I mean them as directly as that. It is as if my name for any actual object were a pointer, and by coming to the object I followed up the pointer and exhibited that to which it already pointed, thus bringing out that it was really that and not anything intermediate that was pointed to. And what I now point to is not an *example* of what I mean as I can point to an example of some generalization that I have made. It *is* what I mean. (This might be used to give a sense – though not Hume's – to Hume's contention that the idea of a thing's existence is none other than the idea of the thing. For in a way this seems absurd. But the thought of an actual thing is not the thought of a possibility with the note of actuality somehow added, or of a possibility to which in fact something actual corresponds.)

The name or thought of something past seems to point to its object in just the same way as the name or thought of any other actual thing; yet how can it, since its object does not exist?

Kant's idea that it is only through the representation of time that the possibility of change is noncontradictory is similar to that of Parmenides. But Parmenides rejects the concept of time as nonsensical precisely because it introduces contradiction. Let us imagine that someone is taught (1) to say "red" when a red light is switched on before him, "yellow" for a yellow light, and so on; and (2) next to say "red", "yellow", etc., when lights of the ap-

---

From Max Black (ed.), *Philosophical Analyiis* (Ithaca, New York, 1950).

propriate colours *have* been switched on but are now off. Imagine a spectator who finds (2) unintelligible, since the learner is not corrected for saying "red" when there is no red light, but this is accepted. It is as if the learner were taught first to act according to a certain rule and then to break that rule. The incompatibility is only removed by our introducing the ideas of the present and the past into our formulation of the rule according to which he is taught to act: "He is taught to say 'is red' when a red light is showing, and 'was red' when a red light not is but has been, showing." – But he was not taught to say "is red" or "was red", but only to say "red", first as described in (1), then as described in (2). But what he does in (2) *contradicts* what he learns in (1) unless we suppose that "red" in (2) really means "was red". There is, of course, no reason to think it impossible that he should behave as is described – one can act quite differently at different times. But then to learn the procedure of (2) is to unlearn the procedure of (1) – and this we do not want to say. We want to say that he learns something in (1) which he applies in (2). But just what he learns in (1) – namely to utter the names of the colours in face of the appropriate lights – is annulled in (2). He does indeed repeat the names of the colours, but he does not apply them as he did. Yet we want to say that he does apply them as he did, namely to the same colours, only now not when they *are* but when they *have been* showing. But what is the explanation of this "have been" which is introduced to explain that the learner is still using, and not misusing, the names he has learnt? One points to red and says "The explanation is that what *was* showing *was* just the same as this"; nevertheless one tends to concentrate on "the same" and "this" without noticing the problem raised by "was". To bring out this problem, let us imagine that we teach "the same" and "this", and that our learner not only learns the colour names but also learns to say "the same as this" correctly for various colour samples presented to him. Now has he everything he needs in order to say "*was* the same as this", without any *new* training? Or, to put it differently, is it unimaginable that he should have got so far but should be unresponsive to attempts to teach him the procedure of (2): is it not, rather, natural that he should hesitate at (2) as if he were being tempted to do what he will be corrected for? But now suppose that he does respond and uses the colour names as described in (2). The explanation "'red' in (2) really means 'was red'", or "refers to the red that *was* showing" would only be an explanation for someone who understood "was red" but had failed to see that this was the rule according to which the learner was being taught to act. But we are not concerned here with such a misunderstanding, but rather with what is understood by someone who does understand the explanation. "Red" in (2), or "was red", is a new use of "red", whose outstanding feature is that "red" can be said when there is no red. It is of no use to say to someone philosophically perplexed at this that "red" is said when there has been red. For he is looking for, and cannot find, a difference between there having been red and there not having been red. We say "He is looking in the wrong place, in the present and not in the past". But what is it to look in the

right place? He is looking for a justification; for surely "was red" is not said without justification when it is said rightly. But there seems to be no justification unless he finds himself reiterating the very thing he is trying to justify.

It seems possible to show someone what to mean when one wants him to say "red" with meaning, but impossible to show him what to mean by "was red"; for how does one get his attention directed to what he is to speak of? When one has to teach "red" one can at least ensure that the learner's eyes are looking in the right direction; and one would not expect to be able to teach him except on this condition. But if one is trying to teach the use of the past tense, then there is nothing to which one can direct his attention and nothing in him to direct in the hope of directing the attention, as in the other case it was possible to direct the eyes. Yet it seems that a necessary condition of his being able to grasp the meaning of "there was red" *is* his attending to the right thing, i.e. to the past showing of red, and that this can only happen if his mind is looking in the right direction: but how can his mind look in the right direction unless he already has the idea of the past? It might seem enough to *have* directed his attention to the red light, which one now recalls by using the expression "was red". But that he *did* attend to the red light does not mean that he *now* has an experience for which he can learn the expression "was red". Just as the occurrence of a succession of ideas does not provide an adequate explanation of the idea of succession, so the fact of a past experience does not provide an adequate explanation of the idea of the past.

It seems that memory must be the key to the problem; that it is memory that gives one the essential meaning of the past tense; that it is like an eye that can look in the required direction, or like sight that corresponds to visible space. So that if the learner in saying "red" in (2) remembers the red light and means what he remembers, *then* he says it rightly, just as he says "red" rightly in (1) if he sees the red light and means what he sees. I know what it is to remember, and what memory gives is the past. We speak of past times and events that form no part of our memories. But (the suggestion is) a past event is something that could be remembered, and someone who remembered it would really know it as past; others understand it because the grammatical form used in describing it has been given sense for them by their own memories. But is it memory as knowledge or as experience from which I am supposing that I derive the idea of the past? As an experience, apparent does not differ from genuine memory; and an apparent memory may be false, or even if it is true it is not genuine memory unless the person who has it was a witness to what he remembers. This shows that I could not use the idea of memory as knowledge – i.e. memory that is both genuine and true – for an 'analysis' of the idea of the past, since my analysis would implicitly include unanalysed references to the past. Still, without intending any such analysis, and without pretending to state any conditions under which memory was knowledge, I may take examples of memory which I insist *are* knowledge.

I think of something that I remember, such as that I had coffee for breakfast, and I want anyone, with whom I am discussing, himself to recall

something that he can remember with certainty. (I am not arguing from the *fact* that I had coffee for breakfast, for, unless he was with me and can remember it too, I do not expect it to be his example; the point is for him to have the idea of finding an example.) Now I say, "This and similar cases must show you the meaning of the past tense. For are you not certain of them; do you not want to call them knowledge? But you cannot know without understanding what you know. Since then, you know this, the meaning of the past tense is contained in what you know. Scrutinize what you know and you will see that meaning in it." Suppose that I am thinking of a representation, *B*, of a certain fact, *A*, and that I want to understand the meaning of *B*. I insist that I know *A*, and that I understand *B* by understanding it as the representation of *A*. If I ask how I know *A*, or how I understand what it is for *A* to be the case, I answer that I simply do know it and understand it: and this I cannot explain further. And it is this knowledge and understanding that explains *B*. Further questions are like raising a doubt about *A*, and, as I am certain of *A*, I reject them.

But how can I explain *B* to myself as a representation of *A* unless I have some other representation of *A* which I use in explaining *B*? If I am confronted with *A* itself, I may think I can evade this difficulty somehow. But if I am not and cannot be confronted with *A*, because it is past, then it is futile for me to appeal to *A* as explaining *B*; for I can only appeal to *A* by using a representation, and the question arises anew about this representation, whatever it is. Knowledge of a past fact can only explain the meaning of some representation of that fact to someone who already has some *other* representation of it; i.e. such explanation is trivial and does nothing for us philosophically. Thus the insistence that I *know* something that has happened is philosophically irrelevant; nor do I want to raise a doubt about such knowledge. To insist that I *know* that such and such has happened is to protest that it *has* happened: if the question is what it means to say that it has happened, one is not helped to answer the question by repeating that it certainly has happened.

It is different if I think that the *experience* of memory is what gives me the idea of the past, and say that memory need be neither genuine nor true in order to contain this idea. Both a false and an apparent memory are of what *seems* to have happened. If I say this, I am holding that memory gives me the form of representation of the past by supplying me with representations which all have this form and that the distinctions of apparent and genuine, and of false and true, are posterior. For a *sense* occurs in a false representation no less than in a true. If, then, I regard my understanding of the meaning of the past tense as got from my memory representations and regard these as the fundamental data for my understanding, then I regard the true and the false indifferently.

Or: I ask myself how I know what the past tense means. And now I say "I remember . . ." thus giving myself an example of a statement about the past, which, I argue, has sense for me by being an expression of a present mental

content. I am on secure ground, because what I say is not doubtful even though what I remember did not happen or I was not there. If I base my understanding on this, then it is secure. I cannot be challenged on the ground that I have covertly introduced an unexplained understanding of the past tense into my pretended explanation of what I understand by it: for all I claim to understand is my own act of recollection, and *that* I understand by making it; and I do not claim for it that it is true, or even that it is genuine, memory.

It is natural in attempting to give an account of memory to draw an analogy between memory and perception, by introducing the notion of memory-data, whose place in the theory of memory is made to correspond to that of sense-data in the theory of perception. When we perceive anything, our senses receive certain impressions, and one of the main philosophical questions about perception is whether we should regard these impressions as "data", as the material for judgements of perception. One can ask someone who is perceiving something to concentrate on the sensible appearances of what he is perceiving, or on the impressions that he is receiving, and to describe these. The following example will bring out the doubtfulness of the analogy between memory and perception that is suggested by the idea of memory-data. Suppose that I witness the commission of a crime and that afterwards I say I remember seeing a certain man standing at the scene of the crime. Someone who does not believe that I have told a lie or made a mistake of memory may nevertheless believe that that man was not present. He discovers a waxwork figure which resembles the man, arranges it in the place where I say I saw him, shows it to me, and I say "Yes! That is exactly what I saw", or "Move it a little this way and draw that curtain. Now it is exactly what I saw". I now become convinced that my previous memory judgement was in fact false, and yet even if it was false, it was not a mistake of *memory*. The explanation of this might seem to be that the mistake was made at the time when I witnessed the crime, and was therefore not a mistake of memory. But suppose – as is possible – that at the time I did not think "That is so-and-so," and took no notice of what I saw, but afterwards said "I saw so-and-so". In that case the mistake was not made at the time of seeing but apparently only at the time of remembering, yet it was not a mistake of memory. But *would* I not have thought, at the time, *if* I had taken account of what I saw, "That's so-and-so"? How is it that I can tell what I would have thought? When I say "I would have thought that if I had taken account of it", I am not necessarily denying that explicit consideration on the spot might have made me think something quite different: "I would have thought it" is a way of expressing, rather than of proving, that the mistake was not one of memory. Is it, then, that I remember a state of mind, which was not an articulate thought, but the not-yet-consciously-articulated thought "That is so-and-so"? Then is this what I *really* remember? Yet when I said "I saw so-and-so" I need not have been thinking about what my state of mind was at the time. Or am I to say that what I really remembered was the

appearance, which I misinterpreted as the appearance of the man when it was really the appearance of the waxwork? Yet it might be that I did not remember the appearance until I was asked to recall it, or until I was confronted with it again. And when I did remember it, what I had was simply a new memory, which led me to reject the old one. One can ask someone who remembers something to consider just what appearance he saw. But this is not drawing his attention to a memory-datum; it is trying to elicit a new memory – of a sense-datum. If, then, I still wish to use the idea of a memory-datum, I must try to say what it is. In the case that I have described, for example, is it just the memory *that* I saw so-and-so? But it is strange to call this a datum, for the memory is a falsehood, so it should seem to follow that the datum is a hallucination or what is called "a nonveridical appearance"; but these expressions do not seem to apply. The memory is, of course, a judgement; but if we ask whether the judgement expresses the possession of a datum or is less immediate than that, the question does not seem to apply – unless we think of the datum in question as belonging to the original occasion; but this is not what is meant by the idea of a memory-datum. If we look for a datum, then should we ask the rememberer to consider exactly what he is experiencing when he has the memory (as we can ask "Just what impressions are you receiving now when you say you see a chair?")? But he may not be able to produce anything that seems relevant if he has no image, and it is not necessary that he should have an image. It is, of course, possible to speak of memory-data in the sense "remembered data", or to speak of memories as data ("My only data for reconstructing this scene were my father's memories"); but it is a mistake to say that memories are constituted by a particular kind of data together with judgements founded on them. The idea of a datum arises originally in the discussion of philosophical questions about sense perception, where there is a contrast between descriptions of impressions and of objects, both of which are noninferential. There would be a point in speaking of memory-data if there were two sorts of memory judgement, the memory-impression judgement and the description of the remembered things. If one makes the analogy between memory and sensation that I have been discussing, one is led to assume that memory is always accompanied by memory images. For it is difficult to see what a memory-datum could be except a memory image. But in fact it is not necessary that a memory judgement should be accompanied by images.

If we reject the analysis of memory in terms of memory-data, it appears less plausible to think that the idea of the past is derived from memory contents, or that it characterizes them, in such a way that it can be 'seen in them'.

The examples that I need to consider in order to clarify the problem are those in which something is (ostensibly) remembered, but there is no memory image. For example, someone asks me what I had for breakfast, and I say "I had coffee": he then asks me, "Did you have an image when you said 'I had coffee'?" and I say "No". Yet I (ostensibly) remembered that I had coffee for breakfast. If I consider what took place in me, I say that I had a memory or at least an apparent memory that I had coffee for breakfast; and I can think of no further mental occurence that is in any way relevant or in which I can say this memory consisted. If I want to maintain that it is the experience of remembering that shows me the meaning of statements about the past, then I ought to be able to maintain that I can see that meaning in the experience of memory that I have in this example. But what I have got is the experience of seeming to know as a witness, without inference and without being told it, the truth of a statement framed in the past tense. I shall therefore be arguing that I understand the past tense because I know what it is to judge that something has happened, noninferentially and as a witness. But the same objection applies to this as applies to the idea that the *knowledge* that something has happened shows one what the past tense means. If the question is "What does it mean to say that such and such happened?" one is not helped to answer it by saying "It *did* happen"; nor is one helped by saying "I have the idea of its having happened, without being told and as a witness". For the question is "What is the idea?" If I could think that a memory judgement was necessarily made in face of a memory-datum distinct from it, I could seek in the datum what would show the sense of the judgement or would supply some foundation for it. Now, however, I am considering an example in which the ostensible memory is made such by being (at least ostensibly) the judgement of a witness about the past. If therefore I say that the experience of memory shows me what the past tense means, this amounts to saying that the experience of seeming to know something about the past as a witness shows me what the past tense means. Thus in appealing to memory I am not appealing to something which I can understand independently of understanding the meaning of the past tense.

I was tempted to use the idea of memory to explain how the learner who named the coloured lights after they had been switched off was able to mean them. If he said "red" etc. *remembering*, then his behaviour was intelligible and intelligent; his use of "red" in (2) was not a misuse; he was applying the colour names to their appropriate objects as he learned to do in (1), only he was now applying them to past, not present objects. But now it appears that I cannot understand what it means to say he said "red" *remembering*, without understanding what it means to say that he meant the past event; hence my explanation is not an explanation but contains within itself the thing that it purported to explain; nor can I think that if I say he remembers I am saying that he has something that *shows* him the meaning of the past tense. I am saying that the learner acted intelligently if his saying "red" in (2) was an expression of knowledge of the past showing of the red light; I am not stating an independently intelligible *condition* on which his utterance would express such knowledge.

I now seem to be brought to the position of saying that *I* know what it is to have (or apparently have) such knowledge, and that it is only if that understanding which *I* have is in the learner that he can mean "red" in (2) as a

report of the past showing of the red light. Suppose there to be in me a state of consciousness, an experience, which I call a memory. What makes this a memory of something that has happened? The memory is an idea that such and such happened. I will suppose that the thing I say happened did happen, that I did witness it; and now I have this experience. The experience could surely be the same even though the idea were inaccurate and though I was not a witness. Then what makes my state or act of consciousness a memory *of that thing*? Is it the mere fact that the thing happened and that I witnessed it? In that case there is nothing in the memory itself that makes it refer to the actual past event. And if so, why should the experience of memory have anything to do with actual past events or show one what it means for something to have happened? If the expression "it happened" is made intelligible for its user by being the expression of a state of consciousness, it cannot be understood as anything but the expression of a state of consciousness. The experience will make one understand what it is to have this state of consciousness. But what has that state to do with *the facts*? If I appeal to the experience of memory to show me what is meant by the past tense, then all I can claim to understand is what it is to have an *idea* which for some reason I call the *idea* of something's having been so: I cannot claim to understand what it is for something to *have* been so. To make this clearer, let me consider something that is taking place before me (or within me) now. Later I shall be able to speak of it. How will what I say be related to the event? How can the attribution of anything to the state of mind I shall *then* be in establish any connection between my words *then* and this thing that is happening *now*? When I look into the past and wonder how I can mean any past thing by what I say – how my thoughts can reach the past thing – I cannot get at that thing except always by means of representations; so my enquiry how my words represent the actual events may seem an impossible one. How can I conceive an event-in-itself beyond the representations, and so ask for a connection between the representations and the event-in-itself? But if I consider some present thing (which can, if you like, be a state of mind) and my future ability to speak of it, it is brought out more clearly how difficult it is to make out that anything I may attribute to my future mental state will make what I say refer to *this*. My mental state may have any characteristics I please to imagine; but if it could be imagined to have those characteristics without being veridical it will have no relation to *this* event except that of coming after it, or possibly of being caused by it. And these relations do not seem to be enough to make it a memory of this event.

In general we must fail if we try to explain the sense of statements about the past by means of present memory, consciousness of meaning, quality of images, or anything else of the kind. For either we have left out all reference to the actual past, or we have surreptitiously introduced it into an explanation that proposed to do without it.

When one thinks of the meaning of a statement about the past, one often thinks of saying, "If I (or anyone else) had been there then, we should have seen, heard . . ." and so on. Thus Mr R. Rhees says that there is not any special difficulty (for a phenomenalist) in explaining what it means to say that material things existed before there were living beings:

> If there are difficulties, they apply to the phenomenalist interpretation of *any* statement about what is not immediately observed. . . . The phenomenalist would say you have still to give an account . . . in terms of the observations that could have been made. . . . The question of how or whether human beings could have got into positions to make these observations at these times is not really relevant.[1]

The relevant question, however, is not how human beings could have got into positions to make those observations at those times, but how it is that we understand "those times". We proceed as if we had a prepared empty time scheme and the only work to be done were the analysis of what is to be put into the various places of the scheme. If someone tells me the criteria, say, for a person's being dead, I do not ask him for what date these criteria hold good: they are the criteria at any date, or rather date does not come in. The description of the criteria tells us in what circumstances we are to say a thing; we think the actual situation in which the criteria occur supplies us with, so to speak, a sense-giving apparatus for the words. But we use the words although the apparatus is not present, and now when we think of criteria we think of the *possibility* of their application. So we hear it discussed what are the criteria for there being a table before one; and then "There is a table in the next room" is explained by means of such remarks as "If one were to go into the next room one would see . . ." and so on. Now suppose a question is raised about the meaning of such hypothetical statements: a great difference can be seen between "if *p* were to happen, *q* would happen" and "if *p* had happened, *q* would have happened". For in "if *p* were to happen, *q* would happen" the antecedent is fulfillable. The meaning of this "if . . . then" can be shown by means of actions or by means of what happens, where the antecedent is fulfillable. One makes a conditional threat and carries it out if the antecedent is fulfilled; one makes a conditional prediction and remarks that one was right if the consequence follows, or that one was wrong if it does not follow; and if the antecedent was not actualized, one does not thereupon claim to be either refuted or confirmed. The sense of the conditional is learnt from such examples. But the sense of the unfulfillable past conditional could not be made clear in any such way; it is difficult to imagine how it could make sense to someone who did not already understand both the fulfillable conditional and the use of words to describe the past – that very past time that is spoken of in the conditional, not some other time that, e.g., falls within his memories. If we have described the criteria for a certain happening by saying what observations justify the statement that it is taking place, then it does follow that if there was such a happening in a certain place and time, some such observations could have been made at that place and time; but this is a

[1] *Mind*, LVI (1947), 380.

mechanical drawing of consequences and is of no use as a contribution to discussion of the meaning of statements about the past.

I am at this point tempted to say: a statement about the past cannot possibly have present criteria, it can only have present evidence. For if I think that a statement about the past has present criteria, must I not suppose that it is possible for the past to change; will not a change in the things that serve as criteria involve a change in the truth of the statement for which they are criteria? But nothing that happens now could change what has happened: this would be an absurdity except in a particular kind of case in which the description of an event depends on what happens after it (e.g. Aristotle describes the use of the Greek word which we translate "involuntary" in such a way that an action can come to have been 'involuntary' if it produces consequences which the agent regrets).

"The past cannot change." It might be retorted that the past has constantly changed: first one thing has happened, then something quite different. But a denial of this is not what is meant by "The past cannot change". Things have taken a certain course, which perhaps can and perhaps cannot be reversed; some actions can be undone. But it makes sense to wish that they had never been done, and when one says "The past cannot change" one is stating that this wish is unfulfillable. But "a change in the past" is *nonsense*, as can be seen from the fact that if a change occurs we can ask for its date. If the idea of a change in the past made sense, we could ask the question "When was the battle of Hastings in 1066?" and that not in the sense "When in 1066 was the battle of Hastings?" The idea of a change in the past involves the idea of a date's being dated. This could of course be given a sense in a particular context – say of a change in a system of dating. But until one gives it sense in some such way – which does not serve one's purpose – it is nonsensical. This consideration helps to remove the impression that when one says "the past cannot change" one is saying of something *intelligible* that it is an impossibility.

If someone said: "The future cannot change, for if it is true that something is going to happen then it is going to happen", this would seem perfectly empty. Similarly "If it is true that something has happened, then it has happened" is empty. "But", it may be said, "there is a difference between the past and future which shows what is meant by 'the past cannot change'. If something is going to happen, it is going to happen; but a change does come about precisely when it does happen. When it has happened it is no longer going to happen. Thus the future changes in just that sense in which we say that the past can't change." To say this is to regard the happening of an event as an irreversible change that takes place in it, and fixes its character. These are two different points, which I might represent in the following way. Suppose that I have a set of pictures in a row: those on my left represent the past, and those on my right represent the future, and the row moves constantly to the left. The idea that happening is an irreversible change which takes place in the event could be represented by the fact that once a picture has passed me it cannot be removed from the row, whereas a picture on the right-hand side can be removed. The idea that happening fixes the character of the event could be represented by the fact that the pictures become set as they pass me so that they cannot be altered whereas the pictures to the right are in a fluid state or are as yet mere blanks. This is an exact image of what I have in mind when I speak of the changelessness of the past. In the image the impossibility of removing or altering the left-hand pictures is a fact which could be otherwise. I use the image to represent what I regard as a fact which could *not* be otherwise. But I have to admit that the force of the idea that I am stating some kind of fact when I say that the past cannot change lies in the inclination I have to draw such an analogy. In the analogy, the impossibility is empirical. I retain the analogy but say that in the original the impossibility is nonempirical. But if I dismiss such analogies, I am unable to say what is left.

This idea of the past as something *there*, to which true statements about the past correspond as a description corresponds to the object that we can compare with it, is what produces the puzzlement of which this paper is a discussion; for now when I wish, as it were, to locate this object I cannot do so. My thought of past things seems like a pointer that points to nothing: and yet to say that the whole conception is a mistake seems to be like denying the reality of the past. This idea, of course, is not what anyone could put forward as an *account* of the meaning of statements about the past. The situation is rather that one has this idea and then tries to justify it, or to find out whether there is any sense in it; whether it contains a point which there is some acceptable way of expressing, or whether it is possible to get rid of it without thinking that one has let go of something essential to an understanding of the past tense.

It is clear that it does not make sense to say that the past has changed. But this does not make it clear that it is equally nonsensical to deny the possibility of a change in the past; it may seem that the idea of a change in the past is nonsensical *because* this changelessness is so absolute a necessity. Nor does it solve the problem to say that the necessity is one of meaning and not of fact, for then the question arises what such a necessity of meaning is: is it that we have an insight into meaning that is expressed in the idea of the changelessness of the past, and is it the business of philosophy to bring about this insight, to distinguish it from recognition of empirical facts, and to leave it at that? How is it that one can, as it were, see a meaning that is no meaning in the idea of the past changing?

Suppose that a child wanted a cake that it had eaten. That it cannot have it again is a mere physical fact. But suppose that it wanted a bang that it had heard, that is, that actual individual bang. (I will disregard the difficulties of making this supposition.) How could it even express such a want? "Make that bang again!" does not express what we want expressed. We can devise a highly artificial[2] method of doing so. We know how to use proper names. Let

[2] This artificiality makes the problem seem much easier to solve than it actually is.

us name a particular bang and call it "*A*". Now we are able to express the desire: "I want *A* again". And we say that it is impossible to get what is asked for. If *A* were a cake which had been eaten, the impossibility of getting it again would be of a different kind. One has the idea that this is because of the nature of a bang and the nature of a cake, and that it is this that we see when we see that we cannot have *A* again if "*A*" is the proper name of a bang. It is as if in prescribing that "*A*" should be the proper name of a bang we were directing the attention to something which we see cannot be repeated: this impossibility is a consequence of its nature, not of its physical but of its logical nature. It is a consequence of the physical nature of the cake and the process of digestion together with certain laws of nature that it cannot be got again after it has been eaten. But – and it is important that we want to stress not the analogy but the contrast – it is a consequence of the *logical* nature of the bang that we cannot have it again. Wittgenstein[5] argues that in such examples, while we think we are insisting on the contrast, we are misunderstanding the one case on the analogy of the other: but in fact what we do with the name – i.e. that we do *not* speak of getting *A* again – is a *part* of (not a consequence of) its being the proper name of a bang. It goes to characterize the use which makes "*A*" have that meaning. If a bang were made in response to this request and satisfied it, then this would show that "*A*" was not being used as the proper name of a bang. For we could not say that the asker was satisfied because he mistakenly supposed the new bang to be the old one again. What sense could there be in the idea (that we are supposing to exist) of the numerical identity of a bang, if a bang heard now could be identified – even wrongly – with a bang heard before? That "*A*" is the proper name of a bang means that we do not speak of getting *A* again. "Getting *A* again" is an expression similar to ones which have use in other contexts, as when "*A*" is the name of a cake. When we transfer it to this context we do not transfer its use; for to describe its use we should have to describe in what circumstance we should say we had got *A* again, as we could do if "*A*" were the name of a cake. But though we do not transfer its use we think we transfer some meaning and so we think that what is meant is something impossible. We think we cannot imagine getting *A* again because of the essential character of what is denoted by the name. But the real reason is that "getting *A* again" is an expression for which we have yet to invent a use in this context; so far no use for it exists. This doesn't seem enough, however: we think we *could* not give it a use meaning that we could not give it the use it has in other contexts, the use that the form of expression suggests or reminds us of. We may stop thinking that the request is impossible because it is a request for the impossible and come to think that it is impossible because it could not make sense; i.e. it is a necessary fact that the use of "*A*" as the proper name of a bang

[5] In this example I have repeated some remarks made by Dr Wittgenstein in discussion. Everywhere in this paper I have imitated his ideas and methods of discussion. The best that I have written is a weak copy of some features of the original, and its value depends only on my capacity to understand and use Dr Wittgenstein's work.

excludes the use of the expression "getting *A* again". And this is because we *could* not describe the circumstances in which we should say we had got *A* again. For if we say merely that we *do* not speak of getting *A* again, this will seem arbitrary or accidental; it might be objected that we *could* do what we do not in fact do, unless it is impossible.

Suppose that someone learns to perform the following exercise, with little connection with anything else he says and does (as might be the case if a child were set to learn dates in the void): he learns to say after his teacher, "The battle of Hastings was in 1066; the battle of Waterloo was in 1815" and so on for a number of battles. Now the word "when" is introduced in this way: the teacher says, "When was the battle of Hastings?" and the learner responds, "1066", and so on for the rest of the battles. This is all that happens; but there is correct and incorrect response, and the learner learns to make only the correct responses. And now if the teacher says, "When was the battle of Hastings in 1066?" the learner has not been taught what to do; what he has been taught to do in response to "when" has not included any response to this question. Suppose that there is nothing that he does spontaneously which is accepted by the teacher. The teacher *could* teach him something to do in response to this new question, but, until he does, the question does not belong to the exercise. This exercise as I have described it has no point. But the *actual* senselessness of "When was the battle of Hastings in 1066?" seems to be like its senselessness in this exercise. In order to give it a sense, one has to invent circumstances, such as a change in the system of dating, in which we are to speak of a date's being dated. The senselessness seems to consist in the fact that we have no use for this combination of words. But it follows from this that the only sense that can be made of the philosophical assertion that the past cannot change is that to speak of a change in the past is to produce an expression for which no use exists and which therefore has *no* sense. For if we could speak of a change in the past, such questions as "When was the battle of Hastings in 1066?" would be applicable. It therefore seems to follow that the *sort* of senselessness in such a question is the *sort* of senselessness that there is in speaking of a change in the past. But the senselessness that there is in that question does seem to be like its senselessness in the artificial exercise that I imagined.

If I consider these examples, I am less inclined to say "The past cannot change" as if I were thereby saying something which could be compared with a statement of physical impossibility. It remains true, nevertheless, that an idea of a change in the past retains an apparent meaning which is one of the sources of perplexity. For this appearance of meaning is such that one wishes to say that one can see that it is somehow not a *legitimate* meaning, and because of this one seems to be saying something positive in saying that the past cannot change. This might be expressed by saying that "a change in the past" is an expression that *could* not be given a sense, meaning that the vague sense that one perceives in it could not be embodied in a use – as if one could understand the sense that it could not be given. It is possible that a reason for

this can be found in the existence of two usages with past tenses; the unfulfillable past conditional sentence and the wish that the past had been otherwise. I have already remarked that the former cannot be imagined to be intelligible except to someone who already understands the conditional and past indicative, and similar remarks, *mutatis mutandis*, would apply to the latter. One can imagine the existence of a people whose language did not include the expression of a wish that things had been otherwise. It would be possible to formulate the wish in their language by using their expression for wishing and their past tense; yet it might be that to them this sounded incomprehensible, or like mere bad grammar. If such a people existed they would seem to be psychologically different from ourselves. The fact that to us this wish makes sense may help to account for the apparent positiveness of "The past cannot change". But if I ask: "Why does the wish make sense to us?" I can find no answer, except that we do use the past tense in this way in connection with wishing.

I considered "The past cannot change" because I was enquiring into the idea of saying "The criteria for statements about the past must lie in the past". Present criteria for past statements seemed to entail the possibility of a change in the past. The reason for thinking this is that if one states the criteria for saying something one may be claiming, or may seem to be claiming, to give a translation or analysis. It would follow that a change in the things that were the criteria for a statement about the past would entail a change in the truth of the statement. But if one gives up the idea that to give the criteria is to give a translation, then this no longer follows. And it is certain that we do use present criteria for statements about the past, and also that no change would make us say "It used to be the case that Brutus killed Caesar, but since such and such a time it has been the case that Caesar killed Brutus". Yet one can imagine changes which would make present criteria for statements about the past inapplicable: constant conflicting shifts in people's memories, chaotic changes in monuments and documents, and so on. And I want to say: Whatever happened – and such things are conceivable – nothing that happened in the future could make it not *still true* that the things that happened in the past did happen as they happened, even though nobody could know it. But this expression, "still true", shows that I have not properly grasped what is implied by saying that the question "When was the Battle of Hastings in 1066?" is senseless. Someone might reply to us, "Yes, Brutus killed Caesar. But how can you make this have any sense but the one given to it by the use which you have learnt for it? How do you think you add anything to that by saying that it would be 'still true' whatever happened?" Yet this is not to say that something could make it false. Still, something could destroy its sense, make it impossible for it to make sense. One wishes to say that the truth is beyond the reach of anything which would destroy the use of the sentence. It is as if one saw *through* the use to the fact to which it is related.

It is possible that I am dominated in this discussion by the idea that in making a philosophical enquiry into the meaning of the past tense I must try to show what it is that we know in knowing anything about the past. Or that it must provide a justification of the concept of the past, as opposed to the normal justifications of particular statements about the past. When I discussed the example of the learner being shown coloured lights, I was looking for something that would make what the learner learned to say stand for the past red light. Suppose it should be objected, "You will not find what you are looking for. You have yourself shown that the *kind* of thing you want is an impossibility. To speak of something past is to have the kind of practice that you have made the learner acquire in your example. The example is, of course, artificial, but a description similar in principle, but far more extensive and complicated, would tell you how words are used in speaking of the past: and to use words in this way *is* to speak of the past." Now for many reasons this criticism is extremely persuasive. In particular the idea of meaning which it indicates is one that resolves the problem presented by the fact that a sense occurs in a false representation no less than in a true one. For this suggests that in an investigation of meaning one ought to attend to the experience of having the idea independently of truth or falsehood; but then one gets into difficulties when one has to relate the idea to the facts. Suppose that one person, *A*, were sitting by another, *B*, who was making reports by telephone to a third, *C*, on events that were taking place before *A* and *B*; *A* is watching and listening, but there are certain words, the meaning of which he does not know, but which he learns on this occasion. (This is shown by the fact that he afterwards uses them correctly.) Now *C* learns the facts; but *A* does not learn the facts, but the sense of the words. Let us suppose that *B* tells lies: *C* is misinformed, but *A* is not misinformed but – if anything – learns a wrong sense for the words. *A* could not learn the right (i.e., the usual) sense of the words in this way except from their being used to make *true* reports. And it is not that *A* sees in what circumstances they *are* true; for he cannot see that a statement is true unless he already knows the sense of the statement. The situation which verifies a remark and that in which the sense of the remark is shown may be identical; and one is strongly inclined to think that in understanding a sense one grasps a fact or an apparent fact. But if one says this, one's problem "What is it to understand this sense?" reappears for the sense of the description of *that* fact or apparent fact.

A particular objection suggests itself to accepting a description of use as an answer to the enquiry how the past tense has meaning. Namely, that this gives the impression that what I know in knowing some past fact is a use. Or rather, since I can hardly be said to know this until I have found the description of it, it seems as if this method of investigation professed to discover that all there really was to know when I knew a past fact was that I was using words in a certain way. But this is in fact a mistake about the purpose of the description, which is not to show one what one is really knowing when one knows something about the past. Indeed, if that were supposed to be its function, it must fail; no one could understand, e.g., the

description that I gave of the practice of the learner in connection with the coloured lights, if he did not *already* understand the past tense; for it was used in the description of what the learner did. The purpose of the description is rather to make us stop asking the question "What is it that I really know?" and stop looking for a foundation for the idea of the past. I spoke earlier of the perplexity which arises from the fact that one looks for a *justification* of "red" in stage (2) of the procedure I described, or of "was red", and cannot find what one is looking for. The purpose of answering the question "How does the past tense have meaning?" by giving a description of use is to make one think that this search for a justification is a mistake.

So far I have spoken only of that use of the past tense in which a witness reports what he has witnessed. We also receive and use testimony, tell and hear stories, make deductions and guesses, use unfulfillable past conditional sentences, express wishes about the past, make historical statements and investigations. If we were to describe the uses of words made in these cases, our descriptions would all lack one particular feature. The descriptions would not in these cases include any mention of actual events corresponding to the past-tense sentences or clauses, such as is made in any description of the personal-report use of the past tense. This would not seem to be of any importance except as regards those uses of the past tense which are supposed to be statements of actual fact. But in these cases one has the idea that a philosophical description of what it is to know the facts necessarily tells one *what* one knows in knowing them, i.e. what the facts themselves consist in. If then the description does not bring in the facts themselves, it seems that what our knowledge really consists in must be considered as reduced to what the description does bring in.

Now it looks as if no theory of knowledge can introduce any mention of actual past events – other than those which are remembered – into its description of what it is to know statements about the past. And so far as I can judge, only the account of meaning given by Wittgenstein enables one without begging the question to introduce mention of actual past events into one's account of knowing the past that one *has* witnessed. This is made possible precisely by that feature of his method which is most difficult to accept: namely, that he attacks the effort at justification, the desire to say: "But one says 'was red' because one knows that the light *was* red!" One says "was red" in these circumstances (not: *recognizing* these circumstances) and that *is* what in this case is called knowing the past fact. To say this is *not* to profess to give an analysis of what one really knows. If we proceed to give similar descriptions of use for statements about the past other than statements of memory and find in these descriptions an answer to the question how those statements have meaning, our descriptions will not include any mention of the actual past events; but as these descriptions will not be analyses or translations we shall not be faced with the difficulty that we have apparently analysed away the actual facts. We shall say, "It is in these circumstances that we speak of knowing such and such; it is this use that gives

that statement a sense." One of the aims of this description of use is to stop one from asking "What do I know?" except as one asks this question in daily life, i.e. unphilosophically; and if one says: "But now it seems to me as if I did not really have any knowledge about the past; all that I can say is that I use words in such and such a way," the reply is that "I have no knowledge about the past" has a sense (e.g. in the mouth of someone who has lost his memory) which is certainly not that according to which one is using it here, and that one has not yet given it a sense as one is now using it.

12 Memory, 'Experience' and Causation

In the *Theaetetus* of Plato, Socrates disputes the definition of knowledge as perception with the objection: If I see, and then shut my eyes and remember what I saw, I still know it, but am not perceiving it. Therefore knowledge is not the same thing as perception. Later, putting a defence for phenomenalism into the mouth of Protagoras, he counters this argument with: "Do you suppose anyone will agree that in remembering what he has experienced, a man is having the same sort of experience (Greek: *pathos*) as he had when he was experiencing what he now remembers?"

The only certain implication of this defence is that memory is not knowledge of the remembered thing. There is a strong suggestion, however, that memory *is* knowledge of the 'experience' that it itself is. No more is said, because Socrates has a weapon to attack phenomenalism more easily. Our interest in *expectation* is concerned with the actual event when it happens (or fails to happen). It would be too patently absurd to construct an analogue to the argument about memory, saying: expectation cannot be the same experience as one has when one observes the expected thing happen, it just is a different one, and therefore expectation cannot be mistaken! Nor does Plato suggest any such argument for Protagoras.

The sentence I quoted about memory is, I think, the first appearance in the literature of the peculiar philosophical concept of 'experience'. In essence this involves an assimilation of memory to sensation. This makes the content of memory present, not past.

Now consider saying, "I don't know if I remember this, or if I was told it". If someone said this about having had a headache that very day, we might look for some extraordinary circumstances to explain his words; lacking any, we should be puzzled to understand him. But when Goethe says at the beginning of his autobiography that he is going to tell of things belonging to his childhood, of which he does not know whether he remembers them or was told them, there is nothing extraordinary about it. It seems familiar and intelligible. Nevertheless, it can raise a puzzle, and it is worth investigating.

In philosophy we sometimes meet, and sometimes may feel inclined to use, the expression "If I remember or ostensibly remember" or "If I at least ostensibly remember". (To the question "ostensibly to whom?" the answer is clearly "to myself".) Now we might formulate the puzzlement that Goethe's remark may elicit by saying: surely he either 'at least ostensibly

From H. D. Lewis (ed.), *Contemporary British Philosophy*, fourth series (London, 1974).

remembered' or definitely did not 'at least ostensibly remember': how could he be in doubt about *that*?

But can I say, can I give expression to an experience by saying, "I at least ostensibly remember"? It would not be at all the same as "I seem to remember". The purpose of this last utterance is rather to say "I remember" somewhat doubtfully, and the theme of doubt is the past, not the present: or, so far as the doubt relates to the present, it concerns one's capacity to tell what happened. Or again, I may say, for example, "Even though I know you didn't have a beard, I keep on remembering you as bearded." But this too does not capture the intent of "I at least ostensibly remember". Nor does "I keep on feeling as if I remembered seeing that happen, though I wasn't there." For if anyone judged that his own memory was false, he *would not say* that he '*ostensibly* remembered'; and correspondingly in the other case. And yet the content of the hypothetical "If I at least ostensibly remember" is supposed to be the hypothesis *merely* that I am experiencing what anyone experiences inasmuch as he is remembering – it being left open whether what I 'at least ostensibly remember' happened or not.

Well, if there is such an experience, let me coin a verb for it. Ought it not to have a direct verbal expression? I will coin the verb 'REMBER' for this.

"I REMBER" will express a *cogitatio*, in Descartes' sense. With 'REMBER' I mean to isolate the central experience of memory from the further implications which are components of the sense of 'remember'; and with "I REMBER" I mean to abstract from any judgement whatever about the past. I then propose as analysis: "He remembers" is a complex idea: "He REMBERS, and the thing happened, and he witnessed it". "I keep on remembering it as it wasn't" will be a complex of "I REMBER that it was . . ." and "it wasn't. . .". "I keep on seeming to remember, though I wasn't there" will be explicable as: "I REMBER, and I wasn't there".

Now, can I fail to know whether I am REMBERING or not? Let us return to Goethe's remark: "I do not know if I remember the things I shall relate, or was told them." Does this mean: "I just don't know if I REMBER this" or something different? Goethe was not in doubt about the things having happened. Is he then implying "I do REMBER this, but I don't know whether the *cause* of my REMBERING this was its having happened in my presence (as, let us suppose, it did) or my having been told?" – If it meant "I just don't know if I REMBER this", then 'REMBERING' is not the kind of verb we tried to make it – for we ought not to be able to say "I just don't know if I REMBER this", any more than we can say "I don't know if I have a headache". But if it means "I do REMBER it but I don't know if that is from my having been there or from my having been told", then the above account of remembering as REMBERING + truth + having been there, and conscious, etc., is inadequate. For it appears that one can still say "I don't know if I remember" though granting that I REMBER, the thing happened, and I was there, etc. What has to be added is the *causal connection*. And how can that get established?

Of course it is possible for us to establish a causal connection between

someone's having an experience and a past event. But in order to do so we need to know how to tell that he has the experience and we must know when something is the expression (or manifestation) of the experience. Now for 'REMBERING' I have given no such criteria. For I coined the name as the name of a supposed 'simple idea of reflection' (in the terminology of the British empiricists) which is meant to be so to speak the core of 'experience' in remembering. I coined the word as a substitute for a circumlocution; and because it *ought* to be coinable if the philosophical implications of that circumlocution are correct. But I explained its sense only by way of the circumlocution. I do not know how we could tell that some simple expression in a natural language had this postulated sense.

The assumption that memory is an experience did make some people think that it consists in the special experience called "having a memory image". This is seldom believed now, because it has often enough been remarked that one *need* not have a memory image when one remembers. But even if somebody has one, and the memory image is just what he finds when he introspects to see what goes on 'when we remember', it *also* needs to be made clear that the memory image is not itself the remembering even on that occasion. One way of seeing this is to suppose the memory image replaced by an actual picture that one sees or draws. Referring an image to the past involves some belief; and it must be done by the person who has the picture before him, if to him it is even a false picture of the past. But in dismissing any image, as *the* experience of memory, we are by no means dismissing the 'experience of memory' itself. If this is a chimerical notion, that still has to be shown.

You may say: a person will at any rate have the experience in question, his expression *will* be the expression of it, if his expression *does* have the sense of "I remember" as this relates, not to 'habitual' memory of information but to 'personal' memory. For in order that one's expression should have this sense of "I remember . . ." there is not any absolute requirement that his memory should actually be true nor that he should actually have witnessed (that is, if he is not lying) what he relates. George IV was not using "I remember" in a wrong sense, just because he said he remembered leading a cavalry charge at Waterloo and he wasn't even there.

That is what has led some thinkers to say that he *was* remembering . Thus Broad:

> In common speech we should be inclined to say that he 'did not really remember' the event in question; just as we are inclined to say the drunkard doesn't 'really see' pink rats. But in both cases this is to mix up psychological, epistemological and ontological considerations in a way which is most detrimental to philosophical discussion. Assuming that the First Gentleman of Europe was correctly describing his state of mind, he was the subject of a situation which has just as good a right to be called a memory situation as a veridical memory of the Duke of Wellington on the same subject-matter.[1]

[1] C. D. Broad, *The Mind and its Place in Nature* (London, 1925), p. 231.

This passage exhibits the very assumption, complete with the assimilation of memory to sensation, which I am engaged in questioning. It is true that George IV will not have been misusing "I remember", or using it in some sense other than the one normal in a purported account 'from memory' of a past action of the speaker. "Remember" in his mouth here *was* the same as "remember" in the supposed statement on Wellington's part. And if the assimilation of memory to sensation were right, that would mean that if he was sincere there was a sense in which what he said was true: he did remember, but what he remembered never happened.

The assimilation is shown false by the following fact: though what George IV thought he remembered need not have happened (in order for him to be using 'remember' correctly), nevertheless he himself could not *leave it open* whether the thing happened or not, whereas in contrast a man can leave it open whether what he sees is purely subjective or is really there. I mean "subjective" in the sense that another could not check his seeing by looking himself. For example, if two people have taken a drug that makes them see white surfaces as coloured, a dispute between them about what colours the surfaces look would be farcical. A man may say what he sees without deciding whether the situation is one of that sort. Rarely, of course. But so is George IV's case rare. The difference between them remains.

Imagine the following dialogues:

(1) "Do you remember whether Jones was here last week?" – "Yes, I remember his being here quite distinctly." – "Ah, well, so *Jones* was here, anyway. Now . . ." – "Wait! I didn't say *that*! I only said I remembered that he was."

(2) "Was Jones here last week?" – "Yes, I remember that he was." – "Oh, someone told you, then? For you were far away yourself." – "No, no one told me, I just remember that he was." – "How do you mean – did you telephone him here, or something?" – "No! Why do you keep asking such silly questions? I only said that I *remembered* that Jones was here. I wasn't here myself, and I didn't learn of it from anything; you know what remembering is like, don't you? *That* is what happened within me when by your question you brought the idea of his having been here before my mind."

Neither of these dialogues makes sense, as they would have to if "I remember" expressed an experience or, as Broad would put it, a 'state of mind' that one is saying one is in.

Let us go back to the remark: "I don't know if I remember this, or if I was told it." If I say this, is there any determining the question? We may certainly be able to determine it negatively. If you were not present on the occasion referred to, then you do not 'personally' remember it, i.e. you do not remember it except in the way you remember any familiar information.

One might wonder how we know this. Is it because we find it never happens? No; the second of the two dialogues above shows the absurdity of even speaking as if it did.

The past tense is heard by the learner of language, and then used by him to

make spontaneous reports. By a report's being "spontaneous" I shall here mean that the person making it did not receive any indications after the event, whether statements or evidence, that the thing happened. Now people do not usually produce spontaneous reports of past events (other than their own past acts, thoughts, sensations, observings, etc.) of which they neither were nor claim to have been *witnesses*. However, they sometimes do. Suppose that we are satisfied that a report is in the defined sense spontaneous, but the person makes no claim to have witnessed the happening – then, if the report is true, we are apt to say he must after all have been there; he 'must be remembering'. This fact alone should tell us that remembering is not (even, so to speak, in its core) quite generally an experience.

If a statement that something happened is spontaneous and true, but we know that the person was not there – what do we say then? According to circumstances, especially what else *he* says, we may say it was a guess (perhaps 'an inspired guess'), or a sudden irrational conviction that happened to be right; or even perhaps that he was a 'seer'. Or we might feel we did not know what to say, and leave the facts at that. But there is one thing we would *not* say. We would not both say that the person was in no way a witness of the previous event, *and* that he was personally remembering that it had happened. That is, "remembering" does not stand for an occurrence which is what it is no matter what else has or has not already been the case, and which therefore we might introduce as an hypothesis in the situation I have sketched. This point serves to characterize the use of any *word* that *means* memory, and can thus justly be called a grammatical point.

But the point seems trivial *if* we can speak of REMBERING. The pure expression for this *would* stand for an occurrence which is what it is no matter what else has been the case. Now if there is such a thing, ought there not to be a 'simple' and primary expression for it? That cannot be "I remember that . . ."; and such an expression as "I remember . . . *as*" is not simple and primary but derivative, secondary. It inherits from "I remember that . . ." the requirement of past presence. Furthermore, we *cannot* just disregard the facts about "I remember that", for remembering is our starting-point, REMBERING has to be something that is supposed always to occur in *remembering*. But now, why should there be any such thing? Remembering is presumably expressed by "I remember". This has to be an expression in a language in which people who have witnessed something make reports of this afterwards. The expression "I remember" may precede such reports. In that case, the hearers are headed off by this precedent expression from bothering about the reporter's present evidence or his other sources of information. *That* will not be in question, but only his reliability first in judging, and later in reporting, things he has been present at. Or if an interrogator, say, does wish to probe into the man's subsequent sources of information, this will be perhaps through suspicion that the witness is not giving a pure memory report of the past. But the question simply does not arise here whether a present memory *experience* took place in someone who

said "I remember", or who simply spoke in witness of what had happened. *That* is not of the slightest concern. It is not what people are concerned with when they ask "Does he *really* remember that?", nor would the absence of such a thing prove one had been wrong to say "I remember". This last consideration shows that it does not save for example Broad's view to say: "Remember" is used in a variety of ways: admittedly we sometimes say "That must have been a case of remembering, though he didn't know it", but "remembering" then does not mean what it means when one says – and *means* – "I remember". The reply to this is: Say, if you like, that "remember" is used differently in "He must have been remembering" ("That must have been a memory, though he didn't know it") and in "I remember". So it is; and we have already given the difference. One man cannot, and the other one can, spontaneously say "I remember". So say if you like: We should really concentrate on the cases where one can spontaneously say "I remember". But even if one does give a privileged position to those cases, *still* it is false to think that 'there must be an experience *of* remembering, which is expressed there'. For in probing the question 'whether he is really *remembering* that, rather than influenced by the talk he has heard or the conclusion he has drawn' we are *not at all* concerned with the presence or absence of such an experience.

As I defined "spontaneous", a report would not be 'spontaneous' if the one who makes it has also heard tell or received other indications of what he himself tells. This is an artificial restriction on the normal sense of "spontaneous". In the normal sense, a report would be 'spontaneous' – relevantly to our present interests – if the person *could* have made it quite apart from any subsequent indication that may have come his way. This comes close to the interest of Goethe's remark. The 'spontaneous' is what the man produces, so to speak, out of the home-baked goods in his store: that is why it seemed natural to introduce causality into the explanation of what Goethe said. It is a metaphor.

For if our account of "I remember" and the past tense is right, this shows that the idea of the causal connection was somehow wrongly introduced. I do not mean that we cannot speak of a causal connection between memory and what is remembered. Quite the contrary. But we introduced it for the wrong purpose and in the wrong place, between the wrong terms. We introduced it in an attempted analysis and to resolve the perplexity aroused by Goethe's remark. On the supposition that the core of memory is an experience – which we are calling "REMBERING" – Goethe would not be in doubt whether he had *that* or not. So, we suggested, he might have meant that, though he did have the experiences, he did not know whether they were caused by his having witnessed the events he was going to recount.

There is a difficulty here, because there is such a thing as being reminded. A reminder elicits a memory. If the core of the memory is an experience, then the reminder will also be a cause of the experience. This seems to throw the whole subject into confusion. One wants to say something like: *that* is not

the sort of 'causing' we had in mind. We want to know the true *parent*, the originating source, of the knowledge; it does not matter what foster parents it may have had.

Noting this difficulty, but boldly disregarding it, let us recapitulate our proposed theory of the meaning of Goethe's remark. First, there is a core in personally remembering an event or situation, which core is an experience. This experience may occur without being a core of remembering. We have called it "REMBERING" and have disconcertingly failed to find any simple primitive term for it in our natural language. Second, REMBERING will be a core of remembering only if REMBERING is caused by having witnessed the events. The theory was: Goethe did have the experience in question but did not know whether it was caused in that way.

Now Goethe also *believed* – indeed he was confident that he knew – that the events had happened. We shall have to add a new component to the situation. We shall have to add that he did not know whether the *belief* was caused by the REMBERING or by something else.

We now find ourselves unclear about what, in this theory, we are identifying as the remembering. Is it the REMBERING, given that this has the right antecedents? Or is it the belief, given that *it* has the right causal antecedents, namely REMBERING with the right causal antecedents?

Now there seems to be no doubt of the following: in such a situation as Goethe describes, a man thinks that such and such occurred. And when he wonders if he knows these things from memory or from being told, he is wondering whether his thinking those things, or his knowledge of those things, *is* memory, not whether it is produced by something else which is a memory. If the supposed intervening experience (given the right provenance) were itself one of remembering that such-and-such occurred, then it would itself already be the judgement that such-and-such occurred: if it were not that judgement, then it would not just have to cause a judgement, but must itself be judged, or judged about: as, if one had an image that struck one as being, or 'felt like', a memory image, one could still not be said to be remembering (even falsely) unless one actually thought, believed, that 'things *were*, for example, as the image represents them'. (Or, again, 'as it does *not* represent them'.) But if one *can* judge "Things *were* as this represents them", why cannot one simply judge "Things were thus and so" – for example, "Jones was there"? We have indeed seen that an image cannot play the part of our supposed experience of REMBERING because beliefs (i.e. memory beliefs) are involved in referring an image to the past. Presumably REMBERING must already contain such a reference, but without being or involving any judgement, any belief. If so, it still has itself to be judged, no less than an image would have to be; though here we can suppose the judgement to be merely *assent*, an inward "Yes, *so*" in face of a present content which already (somehow) has a past reference. And the same consideration applies: if one can say "Yes" to such a content of experience, why can one not equally say "Yes" to the proposition "This happened"? Why must there be an intermediary to have, as it were, the peculiar colour of memory? And if the intermediary is not a required element, neither is causality of it and by it.

Thus the *experience* drops out of the analysis, and the causality – if we are right in introducing it – is between the original witnessing of the event and the present thought, or again, that present state of a human being of which we speak when we say he knows that such-and-such occurred.

We should clear our minds of all prejudice on the subject of causality in considering what we know here. Here is someone, let us say, who knows that such-and-such occurred. Let us suppose that this is a surprising fact – we want an explanation. How *has it come about* that this person knows that? – we ask. And then we are convincingly told: "Well, he was there, he witnessed it." The mystery is removed: we are no longer puzzled about what brought about this state of affairs, which surprised us so much.

This is an original phenomenon of causality: one of its types – whether or not anyone has yet classified it as such.[2] No general theory about what causality is has to be introduced to justify acceptance of it. Nor does it have to be accommodated to any general theory, before it is accepted. It just is one of the things we mean by causality. Only in philosophy our minds tend to be clogged by prejudices on the subject so that we are prevented from seeing the facts that are under our noses.

On the other hand, we do not analyse memory in terms of causality. It is rather that what we call (personal) memory of a past time we also call an effect of the original witnessing of its events. Something is not established as a memory by the discovery of a 'causal connection', nor is ignorance whether something is a memory ignorance of whether it is the product of a known causal mechanism, first activated by an orginal event and remaining set to produce expressions of personal memory that such-and-such happened. (But also: that such-and-such did *not* happen.) We may form such a picture; we may also have a theory that there must be such a mechanism, and perhaps there *is*. But, whether there is such a thing or not, it is not what such a remark as Goethe's is referring to. He is not saying: my present knowledge, which of course (as anyone will agree) is a piece of habitual memory, may originally have been produced by either of two known mechanisms: the personal memory mechanism, or the belief mechanism activated by information; and I do not know which of these bits of machinery did produce it. He is saying only something like this: I am not decisively inclined to say, either that I remember these things happening, or that I do not remember these things happening. I have got this knowledge, and I do not know which sort of knowledge it is. Perhaps I do not know whether I witnessed certain events or not; and even if I did, I do not know whether I could have related them later even if no one had told me of them; nor do I know in every case whether anyone did tell me.

Overmastered by prejudice, someone may be unable to give up the idea of

[2] Russell formulated a notion of 'mnemic causality' in the *Analysis of Mind* (London, 1921), pp. 85–6, which is close to the causality I am speaking of.

the experience of remembering, which occurs at a particular moment, as constituting the essence of what happens when someone personally remembers: it is a definite happening, which one is immediately conscious of; and everyone knows what it is like to have it happen within him. The cases of remembering without realizing that one is doing so do not fit in with this, but no matter: 'remember' is, they may say, a 'family resemblance concept'; they are interested in the central concept, round which cluster some others. Using the central concept, one will say that a man who is not having the experience which, if he gave expression to it, he would express by saying "I remember", *is* not remembering. (Thus Goethe *was* not remembering in this sense, just because he did not know he was.) This experience is necessary for one so much as to have the idea of the past. If such an experience is denied to be a memory on the ground that what it presents as having happened never did, that is like a man's having an hallucinatory sense impression. The hallucinatory sense impression is such because it is not caused by the object it makes one think is there; and similarly for the hallucinatory memory experience.

The dominant idea here is not just that of 'experience': it is that of a definite thing that happens, or a state of consciousness that occurs when (in the 'central' cases) a person *remembers* and which he speaks of when he says "I remember". This idea is common to the 'Cartesian', the British Empiricists, and to the materialist philosophers or proponents of the 'identity theory'. For the latter philosophers say that the mental events or states of mind in question are identical with some brain processes or brain states which neuro-physiology will be able to pick out. They would not say this if they did not unquestioningly accept the same basic assumptions as the 'Cartesian', i.e. if they were not hoodwinked by the suggestions of our verbal forms.

I will show that the role of the supposition of 'causal connection', on such a view, is an unexpected one: that it serves only to rescue the theorist from a form of idealism.

It will of course be necessary that the peculiar experience of memory be quite generally identifiable without connecting it with the past as an actual memory of that. Rather as one could always identify a picture, and indeed as *this* picture, without connecting it with a sitter or an actual scene. The picture, we may say, is a portrait if it has a certain (complicated) causal relation to a sitter. Similarly the memory experience is supposed to be actually a memory if it has a certain causal relation to a past event. But more must be demanded of the memory experience which is to be an actual memory, than of a picture if it is to be a portrait. For the picture does not after all have to resemble the sitter; but the 'content' of the memory experience, if having it is to be an actual case of remembering that . . . , has got to be *true*. The analogy reveals the incorrectness of that idea of the identifiable experience.

For if we explain the picture's being a portrait by its causal relation to a sitter, this will be because we give some causal account of the painter's *intending* this sitter, i.e. intending the picture to be a picture of *this* sitter. But in a memory experience, as we saw before, the 'past' reference, and indeed a quite particular 'past' reference, must already be contained *independently* of the causal relation which we may postulate. That reference would be as integral to the memory experience as the 'content' of the picture is to it; for how otherwise could it be a *memory* experience? It is that thing, which already makes a 'past' reference, about which we are now supposing a causal relation between it and the right past event, to ensure that it is an actual bit of remembering. But it now becomes difficult to see just why a causal relation *is* needed for this. The causal relation certainly does not effect the 'past' reference. I think it is obvious that its role is to secure that the 'past' reference is a real past reference. The 'past' reference that was necessarily involved in the experience is only 'past' in quotation marks. Experience, being present, cannot succeed in actual past reference; the 'past' which is in it is only something present, something to be seen in its own content.

The person believes, that is assents to, the content of this experience. This assent does not, as in our previous argument, involve judgement of what the memory experience presents. For no one has the idea of 'the past' except in the first place from memory, and hence if memory is an experience the idea of the past must simply be in the experience: one cannot bring any primary judgement about the past to bear on the presentation. It is indeed difficult to see what the belief or assent consists in, as it cannot involve assessment of the experience. This was a problem which Hume felt strongly, and could not solve. It seems that assent can only be: allowing the memory experience to feel 'solid'. Plainly we have fallen into idealism. If we are content to remain there, well and good; the causal connection drops out, it is not needed. However we explain 'assent', the causal connection is not needed to do so; and assent to a memory experience which is true is all that is needed for the person to be actually remembering. The distinction of true and false, veridical and non-veridical, will of course not be anything but a distinction between those memory ideas that belong to a coherent picture of the world, and the remaining 'wild' ones.

Thus, if we really adhere to that conception of memory as essentially and centrally a form of experience, the causal connection drops out. This parallels the way in which we saw that the supposed 'experience' dropped out when we tried to give a proper account of Goethe's remark, and did not just dismiss what it describes as a peripheral matter. But the causal connection with a past event outside the experience at once introduces a 'real' past, not yet involved in the memory experience.

I will now return to what I put forward before as the true view: for I consider the idea of the memory experience to have been refuted by my earlier arguments. I argued that there is no question but that there *is* a causal connection between memory and remembered events, but that this is not something discovered: rather, here we are noting a particular type of causality: one of the kinds of things we *call* causation. The causality

which is spoken of here comes in in two ways; one negative, one positive.

Negatively: a person's information about the past may be assigned to a source which *excludes its arising from* his own witnessing of the things in question. If it can be so assigned, then the beliefs he has in the matter will not be personal memories of his. This condition is for the most part satisfied only when his having witnessed something is actually excluded. (Though I imagine an exception to this below.) We must not lose sight of the fact that we sometimes take knowledge or belief as proof that the person is remembering and this in turn as proof of previous witnessing by him.

Positively, the original witnessing of a remembered event is a cause of any present memory (even a false memory) of that event.[5] And this may be seen as 'analytic'. Not so if, instead of speaking of memory, we speak of present knowledge of, or belief about, a past event. An affirmative answer to the question whether the *source* of a man's present knowledge was his having witnessed the event is enough to determine that the knowledge is memory. This is part and parcel of how we use the word "memory". But also, of how we use the word "source" in connection with knowledge.

For being the source and so a cause of knowledge is not the same thing as occurring in some chain of causes or other that leads up temporally to the knowledge. Let us suppose that I witness something and tell others of it: then if I forget all about it and learn of it anew from others who got it directly or indirectly from me, my original witnessing of it will be *a* cause of my present knowledge – but I shall not be remembering.

Perhaps a physiological causal mechanism might be discovered so that in its workings we had a 'right' causal chain: that is, a causal chain which was not detrimental to the present belief's being memory. But even if so, still the existence of a causal chain temporally leading from the original event to the later knowledge does not belong in an analysis of memory.

[5] See Norman Malcolm's first lecture on memory in *Knowledge and Certainty* (Englewood Cliffs, NJ, 1963) for some useful observations on false memory, especially p. 190.

Part Three

Causality and Time

13 Causality and Determination

I

It is often declared or evidently assumed that causality is some kind of necessary connection, or alternatively, that being caused is – non-trivially – instancing some exceptionless generalization saying that such an event always follows such antecedents. Or the two conceptions are combined.

Obviously there can be, and are, a lot of divergent views covered by this account. Any view that it covers nevertheless manifests one particular doctrine or assumption. Namely:

> If an effect occurs in one case and a similar effect does not occur in an apparently similar case, there must be a relevant further difference.

Any radically different account of causation, then, by contrast with which all those diverse views will be as one, will deny this assumption. Such a radically opposing view can grant that often – though it is difficult to say generally when – the assumption of relevant difference is a sound principle of investigation. It may grant that there are necessitating causes, but will refuse to identify causation as such with necessitation. It can grant that there are situations in which, given the initial conditions and no interference, only one result will accord with the laws of nature; but it will not see general reason, in advance of discovery, to suppose that any given course of things has been so determined. So it may grant that in many cases difference of issue can rightly convince us of a relevant difference of circumstances; but it will deny that, quite generally, this *must* be so.

The first view is common to many philosophers of the past. It is also, usually but not always in a neo-Humeian form, the prevailing received opinion throughout the currently busy and productive philosophical schools of the English-speaking world, and also in some of the European and Latin American schools where philosophy is pursued in at all the same sort of way; nor is it confined to these schools. So firmly rooted is it that for many even outside pure philosophy, it routinely determines the meaning of "cause", when consciously used as a theoretical term: witness the terminology of the contrast between 'causal' and 'statistical' laws, which is drawn by writers on physics – writers, note, who would not conceive themselves to be addicts of any philosophic school when they use this language to express that contrast.

The truth of this conception is hardly debated. It is, indeed, a bit of *Weltanschauung*: it helps to form a cast of mind which is characteristic of our whole culture.

Inaugural lecture at Cambridge University (Cambridge, 1971).

The association between causation and necessity is old; it occurs for example in Aristotle's *Metaphysics*: "When the agent and patient meet suitably to their powers, the one acts and the other is acted on OF NECESSITY." Only with 'rational powers' an extra feature is needed to determine the result: "What has a rational power [e.g. medical knowledge, which can kill *or* cure] OF NECESSITY does what it has the power to do and as it has the power, when it has the desire" (Book IX, Chapter V).

Overleaping the centuries, we find it an axiom in Spinoza, "Given a determinate cause, the effect follows OF NECESSITY, and without its cause, no effect follows" (*Ethics*, Book I, Axiom III). And in the English philsopher Hobbes:

> A cause simply, or an entire cause, is the aggregate of all the accidents both of the agents how many soever they be, and of the patients, put together; which when they are supposed to be present, IT CANNOT BE UNDERSTOOD BUT THAT THE EFFECT IS PRODUCED at the same instant; and if any of them be wanting, IT CANNOT BE UNDERSTOOD BUT THAT THE EFFECT IS NOT PRODUCED. (*Elements of Philosophy Concerning Body*, Chapter IX)

It was this last view, where the connection between cause and effect is evidently seen as *logical* connection of some sort, that was overthrown by Hume, the most influential of all philosophers on this subject in the English-speaking and allied schools. For he made us see that, given any particular cause – or 'total causal situation' for that matter – and its effect, there is not in general any contradiction in supposing the one to occur and the other not to occur. That is to say, we'd know what was being described – what it would be like for it to be true – if it were reported for example that a kettle of water was put, and kept, directly on a hot fire, but the water did not heat up.

Were it not for the preceding philosophers who had made causality out as some species of logical connection, one would wonder at this being called a discovery on Hume's part: for vulgar humanity has always been over-willing to believe in miracles and marvels and *lusus naturae*. Mankind at large saw no contradiction, where Hume worked so hard to show the philosophic world – the Republic of Letters – that there was none.

The discovery was thought to be great. But as touching the equation of causality with necessitation, Hume's thinking did nothing against this but curiously reinforced it. For he himself assumed that NECESSARY CONNECTION is an essential part of the idea of the relation of cause and effect (*Treatise of Human Nature*, Book I, Part III, Sections II and VI), and he sought for its nature. He thought this could not be found in the situations, objects or events called "causes" and "effects", but was to be found in the human mind's being determined, by experience of CONSTANT CONJUNCTION, to pass from the sensible impression or memory of one term of the relation to the convinced idea of the other. Thus to say that an event was caused was to say that its occurrence was an instance of some exceptionless generalization connecting such an event with such antecedents as it occurred in. The twist that Hume gave to the topic thus suggested a connection of the notion of

causality with that of deterministic laws – i.e. laws such that always, given initial conditions and the laws, a unique result is determined.

The well-known philosophers who have lived after Hume may have aimed at following him and developing at least some of his ideas, or they may have put up a resistance; but in no case, so far as I know,[1] has the resistance called in question the equation of causality with necessitation.

Kant, roused by learning of Hume's discovery, laboured to establish causality as an *a priori* conception and argued that the objective time order consists "in that order of the manifold of appearance according to which, IN CONFORMITY WITH A RULE, the apprehension of that which happens follows upon the apprehension of that which precedes . . . In conformity with such a rule there must be in that which precedes an event the condition of a rule according to which this event INVARIABLY and NECESSARILY follows" (*Critique of Pure Reason*, Book II, Chapter II, Section III, Second Analogy). Thus Kant tried to give back to causality the character of a *justified* concept which Hume's considerations had taken away from it. Once again the connection between causation and necessity was reinforced. And this has been the general characteristic of those who have sought to oppose Hume's conception of causality. They have always tried to establish the necessitation that they saw in causality: either *a priori*, or somehow out of experience.

Since Mill it has been fairly common to explain causation one way or another in terms of 'necessary' and 'sufficient' conditions. Now "sufficient condition" is a term of art whose users may therefore lay down its meaning as they please. So they are in their rights to rule out the query: "May not the sufficient conditions of an event be present, and the event yet not take place?" For "sufficient condition" is so used that if the sufficient conditions for *X* are there, *X* occurs. But at the same time, the phrase cozens the understanding into not noticing an assumption. For "sufficient condition" sounds like: "enough". And one certainly *can* ask: "May there not be *enough* to have made something happen – and yet it not have happened?"

Russell wrote of the notion of cause, or at any rate of the 'law of causation' (and he seemed to feel the same way about 'cause' itself), that, like the British monarchy, it had been allowed to survive because it had been erroneously thought to do no harm. In a destructive essay of great brilliance he cast doubt on the notion of necessity involved, unless it is explained in terms of universality, and he argued that upon examination the concepts of determination and of invariable succession of like objects upon like turn out to be empty: they do not differentiate between any conceivable course of things and any other. Thus Russell too assumes that necessity or universality is what is in question, and it never occurs to him that there may be any other conception of causality ('The Notion of Cause', in *Mysticism and Logic*).

Now it's not difficult to show it prima facie wrong to associate the notion of cause with necessity or universality in this way. For, it being much easier to

[1] My colleague Ian Hacking has pointed out C. S. Peirce to me as an exception to this generalization.

trace effects back to causes with certainty than to predict effects from causes, we often know a cause without knowing whether there is an exceptionless generalization of the kind envisaged, or whether there is a necessity.

For example, we have found certain diseases to be contagious. If, then, I have had one and only one contact with someone suffering from such a disease, and I get it myself, we suppose I got it from him. But what if, having had the contact, I ask a doctor whether I will get the disease? He will usually only be able to say, "I don't know – maybe you will, maybe not."

But, it is said, knowledge of causes here is partial; doctors seldom even know any of the conditions under which one invariably gets a disease, let alone all the sets of conditions. This comment betrays the assumption that there is such a thing to know. Suppose there is: still, the question whether there is does not have to be settled before we can know what we mean by speaking of the contact as cause of my getting the disease.

All the same, might it not be like this: knowledge of causes is possible without any satisfactory grasp of what is involved in causation? Compare the possibility of wanting clarification of 'valency' or 'long-run frequency', which yet have been handled by chemists and statisticians without such clarification; and valencies and long-run frequencies, whatever the right way of explaining them, have been known. Thus one of the familiar philosophic analyses of causality, or a new one in the same line, may be correct, though knowledge of it is not necessary for knowledge of causes.

There is something to observe here, that lies under our noses. It is little attended to, and yet still so obvious as to seem trite. It is this: causality consists in the derivativeness of an effect from its causes. This is the core, the common feature, of causality in its various kinds. Effects derive from, arise out of, come of, their causes. For example, everyone will grant that physical parenthood is a causal relation. Here the derivation is material, by fission. Now analysis in terms of necessity or universality does not tell us of this derivedness of the effect; rather it forgets about that. For the necessity will be that of laws of nature; through it *we* shall be able to derive knowledge of the effect from knowledge of the cause, or vice versa, but that does not show us the cause as source of the effect. Causation, then, is not to be identified with necessitation.

If *A* comes from *B*, this does not imply that every *A*-like thing comes from some *B*-like thing or set-up or that every *B*-like thing or set-up has an *A*-like thing coming from it; or that given *B*, *A* had to come from it, or that given *A*, there had to be *B* for it to come from. Any of these may be true, but if any is, that will be an additional fact, not comprised in *A*'s coming from *B*. If we take "coming from" in the sense of travel, this is perfectly evident.

"But that's because we can observe travel!" The influential Humeian argument at this point is that we can't similarly observe causality in the individual case (Ibid. Bk I, Pt III, Section II). So the reason why we connect what we call the cause and what we call the effect as we do must lie elsewhere. It must lie in the fact that the succession of the latter upon the former is of a kind regularly observed.

There are two things for me to say about this. First, as to the statement that we can never observe causality in the individual case. Someone who says this is just not going to count anything as 'observation of causality'. This often happens in philosophy; it is argued that 'all we find' is such-and-such, and it turns out that the arguer has excluded from his idea of 'finding' the sort of thing he says we don't 'find'. And when we consider what we are allowed to say we do 'find', we have the right to turn the tables on Hume, and say that neither do we perceive bodies, such as billiard balls, approaching one another. When we 'consider the matter with the utmost attention', we find only an impression of travel made by the successive positions of a round white patch in our visual fields . . . etc. Now a 'Humeian' account of causality has to be given in terms of constant conjunction of physical things, events, etc., not of experiences of them. If, then, it must be allowed that we 'find' bodies in motion, for example, then what theory of perception can justly disallow the perception of a lot of causality? The truthful – though unhelpful – answer to the question: "How did we come by our primary knowledge of causality?" is that in learning to speak we learned the linguistic representation and application of a host of causal concepts. Very many of them were represented by transitive and other verbs of action used in reporting what is observed. Others – a good example is "infect" – form, not observation statements, but rather expressions of causal hypotheses. The word "cause" itself is highly general. How does someone show that he has the concept *cause?* We may wish to say: only by having such a word in his vocabulary. If so, then the manifest possession of the concept presupposes the mastery of much else in language. I mean: the word "cause" can be *added* to a language in which are already represented many causal concepts. A small selection: *scrape, push, wet, carry, eat, burn, knock over, keep off, squash, make* (e.g. noises, paper boats), *hurt*. But if we care to imagine languages in which no special causal concepts are represented, then no description of the use of a word in such languages will be able to present it as meaning *cause*. Nor will it even contain words for natural kinds of stuff, nor yet words equivalent to "body", "wind", or "fire". For learning to use special causal verbs is part and parcel of learning to apply the concepts answering to these and many other substantives. As surely as we learned to call people by name or to report from seeing it that the cat was on the table, we also learned to report from having observed it that someone drank up the milk or that the dog made a funny noise or that things were cut or broken by whatever we saw cut or break them.

(I will mention, only to set on one side, one of the roots of Hume's argument, the implicit appeal to Cartesian scepticism. He confidently challenges us to "produce some instance, wherein the efficacy is plainly discoverable to the mind, and its operations obvious to our consciousness or sensation" (Ibid. Bk I, Pt III, Section XIV). Nothing easier: is cutting, is drinking, is purring not 'efficacy'? But it is true that the apparent perception of such things may be only apparent: we may be deceived by false appearances. Hume presumably wants us to 'produce an instance' in which *efficacy* is

related to sensation as *red* is. It is true that we can't do that; it is not *so* related to sensation. He is also helped, in making his argument that we don't perceive 'efficacy', by his curious belief that "efficacy" means much the same thing as "necessary connection"! But as to the Cartesian-sceptical root of the argument, I will not delay upon it, as my present topic is not the philosophy of perception.)

Second, as to that instancing of a universal generalization, which was supposed to supply what could not be observed in the individual case, the causal relation, the needed examples are none too common. "Motion in one body in all past instances that have fallen under our observation, is follow'd upon impulse by motion in another": so Hume (Ibid. Bk II, Pt III, Section I). But, as is always a danger in making large generalizations, he was thinking only of the cases where we do observe this – billiard balls against free-standing billiard balls in an ordinary situation; not billiard balls against stone walls. Neo-Humeians are more cautious. They realize that if you take a case of cause and effect, and relevantly describe the cause *A* and the effect *B*, and then construct a universal proposition, "Always, given an *A*, a *B* follows" you usually won't get anything true. You have got to describe the absence of circumstances in which an *A* would not cause a *B*. But the task of excluding all such circumstances can't be carried out. There is, I suppose, a vague association in people's minds between the universal propositions which would be examples of the required type of generalizations, and scientific laws. But there is no similarity.

Suppose we were to call propositions giving the properties of substances "laws of nature". Then there will be a law of nature running "The flash-point of such a substance is . . .", and this will be important in explaining why striking matches usually causes them to light. This law of nature has not the form of a generalization running "Always, if a sample of such a substance is raised to such a temperature, it ignites"; nor is it equivalent to such a generalization, but rather to: "If a sample of such a substance is raised to such a temperature and doesn't ignite, there must be a cause of its not doing so." Leaving aside questions connected with the idea of a pure sample, the point here is that 'normal conditions' is quite properly a vague notion. That fact makes generalizations running "Always . . ." merely fraudulent in such cases; it will always be necessary for them to be hedged about with clauses referring to normal conditions; and we may not know in advance whether conditions are normal or not, or what toc ount as an abnormal condition. In exemplar analytical practice, I suspect, it will simply be a relevant condition in which the generalization, "Always if such and such, such and such happens . . .", supplemented with a few obvious conditions that have occurred to the author, turns out to be untrue. Thus the conditional "If it doesn't ignite then there must be some cause" is the better gloss upon the original proposition, for it does not pretend to say specifically, or even disjunctively specifically, what *always* happens. It is probably these facts which make one hesitate to call propositions about the action of substances "laws of nature". The law of inertia, for example, would hardly be glossed: "If a body accelerates without any force acting on it, there must be some cause of its doing so." (Though I wonder what the author of *Principia* himself would have thought of that.) On the other hand just such 'laws' as that about a substance's flash-point are connected with the match's igniting because struck.

Returning to the medical example, medicine is of course not interested in the hopeless task of constructing lists of all the sets of conditions under each of which people always get a certain disease. It is interested in finding what that is special, if anything, is always the case when people get a particular disease; and, given such a cause or condition (or in any case), in finding circumstances in which people don't get the disease, or tend not to. This is connected with medicine's concern first, and last, with things as they happen in the messy and mixed up conditions of life: only between its first and its last concern can it look for what happens unaffected by uncontrolled and inconstant conditions.

II

Yet my argument lies always open to the charge of appealing to ignorance. I must therefore take a different sort of example.

Here is a ball lying on top of some others in a transparent vertical pipe. I know how it got there: it was forcibly ejected with many others out of a certain aperture into the enclosed space above a row of adjacent pipes. The point of the whole construction is to show how a totality of balls so ejected always build up in rough conformity to the same curve. But I am interested in this one ball. Between its ejection and its getting into this pipe, it kept hitting sides, edges, other balls. If I made a film of it I could run it off in slow motion and tell the impact which produced each stage of the journey. Now was the result necessary? We would probably all have said it was in the time when Newton's mechanics was undisputed for truth. It was the impression made on Hume and later philosophers by that mechanics, that gave them so strong a conviction of the iron necessity with which everything happens, the "absolute fate" by which "Every object is determin'd to a certain degree and direction of its motion" (*A Treatise of Human Nature*, Bk II, Pt III, Section I).

Yet no one could have deduced the resting place of the ball – because of the indeterminateness that you get even in the Newtonian mechanics, arising from the finite accuracy of measurements. From exact figures for positions, velocities, directions, spins and masses you might be able to calculate the result as accurately as you chose. But the minutest inexactitudes will multiply up factor by factor, so that in a short time your information is gone. Assuming a given margin of error in your initial figure, you could assign an associated probability to that ball's falling into each of the pipes. If you want the highest probability you assign to be really high, so that you can take it as practical certainty, it will be a problem to reckon how tiny the permitted margins of inaccuracy must be – analogous to the problem: how small a

fraction of a grain of millet must I demand is put on the first square of the chess board, if after doubling up at every square I end up having to pay out only a pound of millet? It would be a figure of such smallness as to have no meaning as a figure for a margin of error.

However, so long as you believed the classical mechanics you might also think there could be no such thing as a figure for a difference that had no meaning. Then you would think that though it was not feasible for us to find the necessary path of the ball because our margins of error are too great, yet there *was* a necessary path, which could be assigned a sufficient probability for firm acceptance of it, by anyone (not one of us) capable of reducing his limits of accuracy in measurement to a sufficiently small compass. Admittedly, so small a compass that he'd be down among the submicroscopic particles and no longer concerned with the measurements, say, of the ball. And now we can say: with certain degrees of smallness we get to a region for which Newton's mechanics is no longer believed.

If the classical mechanics can be used to calculate a certain real result, we may give a sense to, and grant, the 'necessity' of the result, given the antecedents. Here, however, you can't use the mechanics to calculate the result, but at most to give yourself a belief in its necessity. For this to be reasonable the system has got to be acknowledged as true. Not, indeed, that that would be enough; but if so much were secured, then it would be worthwhile to discuss the metaphysics of absolute measures of continuous quantities.

The point needs some labouring precisely because 'the system does apply to such bodies' – that is, to moderately massive balls. After all, it's Newton we use to calculate Sputniks! "The system applies to these bodies" is true only in the sense and to the extent that it yields sufficient results of calculations about these bodies. It does not mean: in respect of these bodies the system is the truth, so that it just doesn't matter that we can't use it to calculate such a result in such a case. I am not saying that a deterministic system involves individual predictability: it evidently does not. But in default of predictability the determinedness declared by the deterministic system has got to be believed because the system itself is believed.

I conclude that we have no ground for calling the path of the ball determined – at least, until it has taken its path – but, it may be objected, is not each stage of its path determined, even though we cannot determine it? My argument has partly relied on loss of information through multiplicity of impacts. But from one impact to the next the path is surely determined, and so the whole path is so after all.

It sounds plausible to say: each stage is determined and so the whole is. But what does "determined" mean? The word is a curious one (with a curious history); in this sort of context it is often used as if it *meant* "caused". Or perhaps "caused" is used as if it meant "determined". But there is at any rate one important difference – a thing hasn't been caused until it has happened; but it may be determined before it happens.

(It is important here to distinguish between being *determined* and being *determinate*. In indeterministic physics there is an apparent failure of both. I am concerned only with the former.)

When we call a result determined we are implicitly relating it to an antecedent range of possibilities and saying that all but one of these is disallowed. What disallows them is not the result itself but something antecedent to the result. The antecedences may be logical or temporal or in the order of knowledge. Of the many – antecedent – possibilities, *now* only one is – antecedently – possible.

Mathematical formulae and human decisions are limiting cases; the former because of the obscurity of the notion of antecedent possibilities, and the latter because decisions can be retrieved.

In a chess-game, the antecedent possibilities are, say, the powers of the pieces. By the rules, a certain position excludes all but one of the various moves that were in that sense antecedently possible. This is logical antecedence. The next move is determined.

In the zygote, sex and eye-colour are already determined. Here the antecedent possibilities are the possibilities for sex and eye-colour for a child; or more narrowly: for a child of these parents. *Now*, given the combination of this ovum and this spermatozoon, all but one of these antecedent possibilities is excluded.

It might be said that anything was determined once it had happened. There is now no possibility open: it *has* taken place! It was in this sense that Aristotle said that past and present were necessary. But this does not concern us: what interests us is *pre*-determination.

Then "each stage of the ball's path is determined" must mean "Upon any impact, there is only one path possible for the ball up to the next impact (and assuming no air currents, etc.)." But what ground could one have for believing this, if one does not believe in some system of which it is a consequence? Consider a steel ball dropping between two pins on a Galton board to hit the pin centred under the gap between them. That it should balance on this pin is not to be expected. It has two possibilities; to go to the right or to the left. If you have a system which forces this on you, you can say: "There has to be a determining factor; otherwise, like Buridan's ass, the ball must balance." But if you have not, then you should say that the ball may be undetermined until it does move to the right or the left. Here the ball had only two significant possibilities and was perhaps unpredetermined between them. This was because it cannot be called determined – no reasonable account can be given of insisting that it is so – within a small range of possibility, actualization within which will lead on to its falling either to the right or to the left. With our flying ball there will also be such a small range of possibility. The further consequences of the path it may take are not tied down to just two significant possibilities, as with one step down the Galton board: the range of further possibility gets wider as we consider the paths it may take. Otherwise, the two cases are similar.

We see that to give content to the idea of something's being determined, we have to have a set of possibilities, which something narrows down to one – before the event.

This accords well with our understanding of part of the dissatisfaction of some physicists with the quantum theory. They did not like the undeterminedness of individual quantum phenomena. Such a physicist might express himself by saying "I believe in causality!" He meant: I believe that the real physical laws and the initial conditions must entail uniqueness of result. Of course, within a range of co-ordinate and mutually exclusive identifiable possible results, only one happens: he means that the result that happens ought to be understood as the only one that was possible before it happened.

Must such a physicist be a 'determinist'? That is, must he believe that the whole universe is a system such that, if its total states at t and t' are thus and so, the laws of nature are such as then to allow only one possibility for its total state at any other time? No. He may not think that the idea of a total state of the universe at a time is one he can do anything with. He may even have no views on the uniqueness of possible results for whatever may be going on in any arbitrary volume of space. For "Our theory should be such that only the actual result was possible for that experiment" doesn't mean "Our theory should have excluded the experiment's being muffed or someone's throwing a boot, so that we didn't get the result", but rather: "Our theory should be such that only this result was possible as *the result of the experiment*." He hates a theory, even if he has to put up with it for the time being, that essentially assigns only probability to a result, essentially allows of a range of possible results, never narrowed down to one until the event itself.

It must be admitted that such dissatisfied physicists very often have been determinists. Witness Schrödinger's account of the 'principle of causality': "The exact physical situation at *any* point P at a given moment t is unambiguously determined by the exact physical situation within a certain surrounding of P at any previous time, say $t - \tau$. If τ is large, that is if that previous time lies far back, it may be necessary to know the previous situation for a wide domain around P" (*Science and Humanism*). Or Einstein's more modest version of a notorious earlier claim: if you knew all about the contents of a sphere of radius 186,000 miles, and knew the laws, you would be able to know for sure what would happen at the centre for the next second. Schrödinger says: *any* point P; and *a* means *any* sphere of that radius. So their view of causality was not that of my hypothetical physicist, who I said may not have views on the uniqueness of possible results for whatever may be going on in any arbitrary volume of space. My physicist restricts his demand for uniqueness of result to situations in which he has got certain processes going in isolation from inconstant external influences, or where they do not matter, as the weather on a planet does not matter for predicting its course round the sun.

The high success of Newton's astronomy was in one way an intellectual disaster: it produced an illusion from which we tend still to suffer. This illusion was created by the circumstance that Newton's mechanics *had a good model in the solar system*. For this gave the impression that we had here an ideal of scientific explanation; whereas the truth was, it was mere obligingness on the part of the solar system, by having had so peaceful a history in recorded time, to provide such a model. For suppose that some planet had at some time erupted with such violence that its shell was propelled rocket-like out of the solar system. Such an event would not have violated Newton's laws; on the contrary, it would have illustrated them. But also it would not have been calculable as the past and future motions of the planets are presently calculated on the assumption that they can be treated as the simple 'bodies' of his mechanics, with no relevant properties but mass, position and velocity and no forces mattering except gravity.

Let us pretend that Newton's laws were still to be accepted without qualification: no reserve in applying them in electrodynamics; no restriction to bodies travelling a good deal slower than light; and no quantum phenomena. Newton's mechanics is a deterministic system; but this does not mean that believing them commits us to determinism. We could say: of course nothing violates those axioms or the laws of the force of gravity. But animals, for example, run about the world in all sorts of paths and no path is dictated for them by those laws, as it is for planets. Thus in relation to the solar system (apart from questions like whether in the past some planet has blown up), the laws are like the rules of an infantile card game: once the cards are dealt we turn them up in turn, and make two piles each, one red, one black; the winner has the biggest pile of red ones. So once the cards are dealt the game is determined, and from any position in it you can derive all others back to the deal and forward to win or draw. But in relation to what happens on and inside a planet the laws are, rather, like the rules of chess; the play is seldom determined, though nobody breaks the rules.[2]

Why this difference? A natural answer is: the mechanics does not give the special laws of all the forces. Not, for example, for thermal, nuclear, electrical, chemical, muscular forces. And now the Newtonian model suggests the picture: given the laws of all the forces, then there is total coverage of what happens and then the whole game of motion is determined; for, by the first law, any acceleration implies a force of some kind, and must not forces have laws? My hypothetical physicist at least would think so; and would demand that they be deterministic. Nevertheless he still does not have to be a 'determinist'; for many forces, unlike gravity, can be switched on and off, are generated, and also shields can be put up against them. It is one thing to hold that in a clear-cut situation – an astronomical or a well-contrived experimental one designed to discover laws – 'the result' should be deter-

[2] I should have made acknowledgements to Gilbert Ryle (*Concept of Mind*, p. 77) for this comparison. But his use of the openness of chess is somewhat ambiguous and is not the same as mine. For the contrast with a closed card game I was indebted to A. J. P. Kenny.

mined: and quite another to say that in the hurly-burly of many crossing contingencies whatever happens next must be determined; or to say that the generation of forces (by human experimental procedures, among other things) is always determined in advance of the generating procedure; or to say that there is always a law of composition, of such a kind that the combined effect of a set of forces is determined in every situation.

Someone who is inclined to say those things, or implicitly to assume them, has almost certainly been affected by the impressive relation between Newton's mechanics and the solar system.

> We remember how it was in mechanics. By knowing the position and velocity of a particle at one single instant, by knowing the acting forces, the whole future path of the particle could be foreseen. In Maxwell's theory, if we know the field at one instant only, we can deduce from the equations of the theory how the whole field will change in space and time. Maxwell's equations enable us to follow the history of the field, just as the mechanical equations enabled us to follow the history of material particles . . . With the help of Newton's laws we can deduce the motion of the earth from the force acting between the sun and the earth.[5]

"By knowing the acting forces" – that must of course include the *future* acting forces, not merely the present ones. And similarly for the equations which enable us to follow the history of the field; a change may be produced by an external influence. In reading both Newton and later writers one is often led to ponder that word "external". Of course, to be given 'the acting forces' is to be given the external forces too and any new forces that may later be introduced into the situation. Thus those first sentences are true, if true, without the special favour of fate, being general truths of mechanics and physics, but the last one is true by favour, by the brute fact that only the force acting between earth and sun matters for the desired deductions.

The concept of necessity, as it is connected with causation, can be explained as follows: a cause *C* is a necessitating cause of an effect E *when* (I mean: on the occasions when) if *C* occurs it is certain to cause *E* unless something prevents it. *C* and *E* are to be understood as general expressions, not singular terms. If 'certainty' should seem too epistemological a notion: a necessitating cause *C* of a given kind of effect *E* is such that it *is* not possible (on the occasion) that *C* should occur and should not cause an *E*, given that there is nothing that prevents an *E* from occurring. A non-necessitating cause is then one that can fail of its effect without the intervention of anything to frustrate it. We may discover *types* of necessitating and non-necessitating cause; e.g. rabies is a necessitating cause of death, because it is not possible for one who has rabies to survive without treatment. We don't have to tie it to the occasion. An example of a non-necessitating cause is mentioned by Feynman: a bomb is connected with a Geiger counter, so that it will go off if the Geiger counter registers a certain reading; whether it will

[5] Albert Einstein and Leopold Infeld, *The Evolution of Physics* (New York, 1938; paperback edn 1967), p. 146.

or not is not determined, for it is so placed near some radioactive material that it may or may not register that reading.

There would be no doubt of the cause of the reading or of the explosion if the bomb did go off. Max Born is one of the people who has been willing to dissociate causality from determinism: he explicates cause and effect in terms of dependence of the effect on the cause. It is not quite clear what 'dependence' is supposed to be, but at least it seems to imply that you would not get the effect without the cause. The trouble about this is that you might – from some other cause. That this effect was produced by this cause does not at all show that it could not, or would not, have been produced by something else in the absence of this cause.

Indeterminism is not a possibility unconsidered by philosophers. C. D. Broad, in his inaugural lecture, given in 1934, described it as a possibility; but added that whatever happened without being determined was accidental. He did not explain what he meant by being accidental; he must have meant more than not being necessary. He may have meant being uncaused; but, if I am right, not being determined does not imply not being caused. Indeed, I should explain indeterminism as the thesis that not all physical effects are necessitated by their causes. But if we think of Feynman's bomb, we get some idea of what is meant by "accidental". It was random: it 'merely happened' that the radio-active material emitted particles in such a way as to activate the Geiger counter enough to set off the bomb. Certainly the motion of the Geiger counter's needle is caused; and the actual emission is caused too; it occurs because there is this mass of radioactive material here. (I have already indicated that, contrary to the opinion of Hume, there are many different sorts of causality.) But all the same the *causation* itself is, one could say, *mere hap*. It is difficult to explain this idea any further.

Broad used the idea to argue that indeterminism, if applied to human action, meant that human actions are 'accidental'. Now he had a picture of choices as being determining causes, analogous to determining physical causes, and of choices in their turn being either determined or accidental. To regard a choice as such – i.e. any case of choice – as a predetermining causal event, now appears as a naive mistake in the philosophy of mind, though that is a story I cannot tell here.

It was natural that when physics went indeterministic, some thinkers should have seized on this indeterminism as being just what was wanted for defending the freedom of the will. They received severe criticism on two counts: one, that this 'mere hap' is the very last thing to be invoked as the physical correlate of 'man's ethical behaviour'; the other, that quantum laws predict statistics of events when situations are repeated; interference with these, by the *will*'s determining individual events which the laws of nature leave undetermined, would be as much a violation of natural law as would have been interference which falsified a deterministic mechanical law.

Ever since Kant it has been a familiar claim among philosophers, that one can believe in both physical determinism and 'ethical' freedom. The recon-

ciliations have always seemed to me to be either so much gobbledegook, or to make the alleged freedom of action quite unreal. My actions are mostly physical movements; if these physical movements are physically predetermined by processes which I do not control, then my freedom is perfectly illusory. The truth of physical indeterminism is thus indispensable if we are to make anything of the claim to freedom. But certainly it is insufficient. The physically undetermined is not thereby 'free'. For freedom at least involves the power of acting according to an idea, and no such thing is ascribed to whatever is the subject (what would be the relevant subject?) of unpredetermination in indeterministic physics. Nevertheless, there is nothing unacceptable about the idea that that 'physical haphazard' should be the only physical correlate of human freedom of action; and perhaps also of the voluntariness and intentionalness in the conduct of other animals which we do not call 'free'. The freedom, intentionalness and voluntariness are not to be analysed as the same thing as, or as produced by, the physical haphazard. Different sorts of pattern altogether are being spoken of when we mention them, from those involved in describing elementary processes of physical causality.

The other objection is, I think, more to the point. Certainly if we have a statistical law, but undetermined individual events, and then enough of these are supposed to be pushed by will in one direction to falsify the statistical law, we have again a supposition that puts will into conflict with natural laws. But it is not at all clear that the same train of minute physical events should have to be the regular correlate of the same action; in fact, that suggestion looks immensely implausible. It is, however, required by the objection.

Let me construct an analogy to illustrate this point. Suppose that we have a large glass box full of millions of extremely minute coloured particles, and the box is constantly shaken. Study of the box and particles leads to statistical laws, including laws for the random generation of small unit patches of uniform colour. Now the box is remarkable for also presenting the following phenomenon: the word "Coca-Cola" formed like a mosaic, can always be read when one looks at one of the sides. It is not always the same shape in the formation of its letters, not always the same size or in the same position, it varies in its colours; but there it always is. It is not at all clear that those statistical laws concerning the random motion of the particles and their formation of small unit patches of colour would have to be supposed violated by the operation of a cause for this phenomenon which did not derive it from the statistical laws.

It has taken the inventions of indeterministic physics to shake the rather common dogmatic conviction that determinism is a presupposition or perhaps a conclusion, of scientific knowledge. Not that that conviction has been very much shaken even so. Of course, the belief that the laws of nature are deterministic has been shaken. But I believe it has often been supposed that this makes little difference to the assumption of macroscopic determinism: as if undeterminedness were always encapsulated in systems whose

internal workings could be described only by statistical laws, but where the total upshot, and in particular the outward effect, was as near as makes no difference always the same. What difference does it make, after all, that the scintillations, whereby my watch dial is luminous, follow only a statistical law – so long as the gross manifest effect is sufficiently guaranteed by the statistical law? Feynman's example of the bomb and Geiger counter smashes this conception; but as far as I can judge it takes time for the lesson to be learned. I find deterministic assumptions more common now among people at large, and among philosophers, than when I was an undergraduate.

The lesson is welcome, but indeterministic physics (if it succeeds in giving the lesson) is only culturally, not logically, required to make the deterministic picture doubtful. For it was always a mere extravagant fancy, encouraged in the 'age of science' by the happy relation of Newtonian mechanics to the solar system. It ought not to have mattered whether the laws of nature were or were not deterministic. For them to be deterministic is for them, together with the description of the situation, to entail unique results in situations defined by certain relevant objects and measures, and where no part is played by inconstant factors external to such definition. If that is right, the laws' being deterministic does not tell us whether 'determinism' is true. It is the total coverage of every motion that happens, that is a fanciful claim. But I do not mean that any motions lie outside the scope of physical laws, or that one cannot say, in any given context, that certain motions would be violations of physical law. Remember the contrast between chess and the infantile card game.

Meanwhile in non-experimental philosophy it is clear enough what are the dogmatic slumbers of the day. It is over and over again assumed that any singular causal proposition implies a universal statement running "Always when this, then that"; often assumed that true singular causal statements are derived from such 'inductively believed' universalities. Examples indeed are recalcitrant, but that does not seem to disturb. Even a philosopher acute enough to be conscious of this, such as Davidson, will say, without offering any reason at all for saying it, that a singular causal statement implies *that there is* such a true universal proposition[4] – though perhaps we can never have knowledge of it. Such a thesis needs some reason for believing it! 'Regularities in nature': that is not a reason. The most neglected of the key topics in this subject are: interference and prevention.

[4] 'Causal Relations', *Journal of Philosophy*, 64 (November 1967).

14 Times, Beginnings and Causes

Philosophical theses sometimes suffer damage from too much success. For reasons of his own the philosopher makes some general diagnosis. As it might be: that all logical truths are tautologies. He creates such conviction that his statement gets to be taken, not as a substantive claim, but as in some way true by definition. Such an over-success happened to Hume's observation that a causal relation is not a matter of logical necessity by which, given the cause, the effect must follow. This got such a hold that it is sometimes argued that some objects (I follow Hume's usage of "object") cannot be causally related, just because there is a logical necessity of the one, given the other. In this way Hume's substantive philosophical thesis has come to be seen as if it had been a partial definition. I don't think that that was his intention. True, he would have rejected counter-examples. But he would have argued against each specifically. Certainly making a definition out of his diagnosis is an uninteresting move.

Now here is a seeming counter-example, at least on widely accepted views. Friction produces heat. What at first sight could be a better illustration of Hume's own thesis? For couldn't you imagine it producing cold instead? But now we are told that heat is a state of increased excitation of molecules. Well, given that the molecules go on existing – and unless they do how can there be friction, for things would crumble away at the mere attempt to rub them together – given that they do go on existing, how then can they fail to be in a state of increased excitation from friction? For that is the rapid motion against one another of two juxtaposed surfaces; hence there will be a mixture of molecules of either object in some places and this must involve a lot of extra banging of molecules upon one another.

So if that account of heat is right, this will be a counter-example to Hume's contention that the ideas of cause and effect are evidently distinct and that "the power by which an object produces another is never discoverable merely from their idea". For the ideas of friction and heat apparently turn out not to be totally "distinct". Are they then not to be *called* an example of cause and effect? That would be the move marking the over-success of Hume's thesis – a stultifying move. Or are we to say that even if heat is increased molecular excitation, the idea of heat is not the idea of increased molecular excitation? – but Hume thought it was the objects that were cause and effect that could always, without absurdity, be supposed to occur one without the other. Here we stumble on the problem of description, which offers us difficulties in formulating Hume's discovery. It is after all a careless use of words to say "causal connections are not logically

From *Proceedings of the British Academy*, 60 (1974).

necessary" because causal connections are between things, events and so on, while logical connections are taken to be between suppositions, or predicates. How then are we to formulate the alleged always non-absurd suppositions? We can always frame descriptions of particular causes and effects, which *are* logically connected.

Hume himself avoids this difficulty. He puts the matter like this: where we have a beginning of existence, and a cause of it, there is no contradiction or absurdity in supposing the one to occur without the other. Thus he continues to speak of suppositions – proper subjects of logical relations – without specifying any class of propositions which in his view are always non-contradictory.

But even though we can state his thesis as a generalization without running into the description problem, we cannot be satisfied to leave it at that. Such a general thesis requires not only particular instances illustrative of it, but also a certainty of its being possible to frame the proposition stating the particular instance of it in every case.

To repeat: we can always frame descriptions of particular causes and effects such that the descriptions are logically linked. For example "Something which caused an explosion occurred" and "An explosion occurred". Now that fact surely cannot damage Hume's thesis! Still, how are we to state the thesis with reference to particular cases? It will be useless to say: *some* statements of the occurrence of the cause and non-occurrence of the effect it produced will be non-contradictory. For that sort of move, were it effective, could be used to show that there are no necessary connections of any sort, no necessary relations between propositions or numbers for example. For, just as we can always form a description of a cause and an effect such that the one logically entails the other, so we can always form accidental descriptions of, let us say, mutually entailing predicates, which do *not* entail one another.

The evasive generalization may be put like this: in general, whatever event E is in question, the cause of E might have occurred and E not have occurred. That is not to say "For every event E this might have been the case: the cause of E occurred and E did not", but rather "concerning the cause of E, it might have been that *that* occurred and E not".

That formulation suggests that there is some description (or: are some descriptions) of the cause, and also of the effect, which are not external, artificial or oblique but which rather simply present the cause, or the effect (in any particular case) in its character just as a happening. Descriptions which convey the physical reality of the thing that happens, and do not get at it by some other means such as "the event mentioned in such a place", or "the (prominent) event which took place in such a room at such a time". "Friction" will be such a "physical description", as I shall call it, and so will "heat".

However, it may very likely be argued that 'Friction produces heat' gives us no falsification of the Humeian principle. For there are both empirical

investigation and theory coming between that bare statement "Friction produces heat", which is the product of the most elementary observation, and the idea that the molecules at the surfaces of the two objects suffer increased excitation when they are rubbed together, which is heat. That is true: but then how are we to understand Hume's thesis? Is it to be taken as saying only that a rudimentary understanding of the objects which are causes and effects will never yield a necessary connection? That so long as we remain ignorant of the nature of things, we will find no logic in the sequence of events? That would not be too impressive. A proper physical statement of the cause will never logically yield a proper physical statement of the effect – that is the thesis. If so, then friction producing heat, on being properly understood, will also be a fair counter-example. It cannot be excluded on the ground that some empirical investigation has gone on before that formulation is reached. For the investigation was into the nature of this cause and this effect.

The description problem is a deep one and I will beg leave to go no further with it here but to assume it solved. Certainly if it can't be solved, Hume's primary discovery was nothing. I believe it was a great one, a great correction of false philosophic assumptions. But also, that it was not true without exception. I do not myself accept the counter-example I have given, because the explanation of heat seems wrong. At very high temperatures the molecular bond is dissolved – so at best some reformulation is necessary. (And think of the radiant heat in the interior of the sun: how absurd to say that *that* is not heat.)

But here is a genuine counter-example. I find an object in some place and ask how it comes to be there. One partial causal explanation would be that it travelled there from somewhere else. Call the place it was at before "*A*", the place where it is now "*B*"; our causal explanation is that it went along some path from *A* to *B*. As opposed, that is, to arriving at *B* from *A* in some other way. And also, as opposed to its coming into existence at *B*. But travelling from *A* to *B* necessarily involves being at *B* – even if only for one point of time.

Suppose our object turned up at *B* after having been at *A*, but without traversing any path from *A* to *B*, then it would be empty indeed to call "It arrived *at B* from *A*" a cause of its being at *B*. "It was at *A* before" might state such a cause, but not "its arriving at *B* from *A*", for that just adds the description of the effect to the description of the cause. Not so "It travelled from *A* as far as *B* by some path". That however entails, not indeed its being at rest at *B*, but at least its being at *B*.

Hume wouldn't accept this as a counter-example. But that is because of his – unacceptable – views on space and time. *He* would have divided the journey into two parts: the part up to immediately before the thing was at *B* and the "contiguous" part in which it was at *B*, and he would have said that at most the first part might be cause of the second. But this requires the absurd notion of the contiguity or strict adjacence of bits of space and time

with no points in common. Thus what I call Hume's major thesis about causality hangs together with his atomism, with his rejection of the continuum. It depends on that for its universality. I count it as not universally true, and would say that Hume showed us only that causal relations do not as such involve logical connections of cause and effect. In fact the example of travel is the only kind of counter-example that I know.

Hume is making a substantive claim, then, not a stipulation about what we are to call "causes and effects". And the claim seems to be almost universally true. I suspect indeed that it can be faulted only where it is tied up with his views on space and time. It is true that causal properties enter into the definition of substances, so that you might think certain effects resulted of necessity – logical necessity – from the interaction of substances. But you would be wrong in making this inference. A *lusus naturae* is always logically possible. A different melting point may indeed prove this isn't phosphorus; the lump of phosphorus turning into a little bird or a piece of bread would not.

Hume's observation ought to have been a very liberating one. But so far few people have been much liberated, for Hume himself, and almost everybody since, has been anxious to forge some substitute for the chains that he broke – to replace the logical necessity by another one just as universal as it.

I want to discuss a quite distinct doctrine of Hume's, which is often accepted with this one in a single package. I mean his doctrine that his form of a "law of universal causation", namely "Every beginning of existence has a cause", has no logical necessity about it.

It is obvious that the doctrines are distinct. Suppose there is a logical necessity about "Every beginning of existence has a cause". It would not follow that the connection between any particular cause and its effect involved any logical necessity. For "Every beginning of existence must have a cause" only says of every beginning of existence that it must have some cause or other, not that there is any given cause that anything must have. Conversely, someone might think that every case of cause and effect, properly stated, involved a logical necessity of that cause's having that effect and/or that effect's having that cause, while not thinking that beginnings of existence had to have causes. (This was McTaggart's view.)

Hume appeared to recognize the distinction of the two questions – but, having once proved to his own satisfaction that "Every beginning of existence has a cause" is neither intuitively nor demonstratively certain (i.e., as people would say now, is not a bit of logical truth), he 'sinks' the question why we believe it in the question why 'we' believe in the necessary connection of particular causes and effects. Well, I don't believe that quite generally, in *any* sense of "necessary connection". We need reason to believe such a thing, and Hume was right in thinking there was none. He was only wrong in thinking that we were under a compulsion to believe it.

Hume's attempted proof that "Every beginning of existence has a cause"

is not a logical certainty is not satisfactory. In *A Treatise of Human Nature* it goes like this.

> As all distinct ideas are separate from each other, and as the ideas of cause and effect are evidently distinct, 'twill be easy for us to conceive any object to be non-existent this moment, and existent the next, without conjoining to it the distinct idea of a cause . . . The separation, therefore, of the idea of a cause from thet of a beginning of existence, is plainly possible for the imagination; and consequently the actual separation of these objects is so far possible, that it implies no contradiction or absurdity. (Book I, Part I, Section III)

The argument is apparently based on a belief that if we can think of one thing (in this case, a beginning of existence) without thinking of another (in this case, a cause) then the one can occur without the other. In its contrapositive form, "If there can't *be* one thing without another, you can't *think* of the first without thinking of the second", Hume certainly assented to this general principle, and in this form its absurdity is now generally realized. It is what leads him to say that you cannot think of anything without thinking of it in full particularity. So it may seem odd that the reasoning here, about causes, is so widely believed. The reason is perhaps partly that what Hume is arguing for here is not always sharply distinguished in people's minds from the great major thesis, and partly also that this argument is credited as an argument from imagination to possibility. "We can imagine a beginning of existence without a cause, therefore there can be a beginning of existence without a cause." But what interest or value can there be in such imagining? I might as well argue that, because I can imagine circle-squaring, that is possible. It is different when I use the argument from imagination in a particular case of cause and effect. For here it may be clear that I do know what it would be like to establish the occurrence of *this* phenomenon without *this* cause.

Having said this, however, *reculons pour mieux sauter*, so that we may find what is true about the argument. Certainly if we look at a thought as a psychological event, and, by the "experimental method" as suggested on Hume's title-page, try what we *can* think without what, the argument lacks all force. But suppose we consider a thought, not as a psychological event, but as the content of a proposition, the common possession of many minds. We may then find a way of putting the point. As Aquinas remarks in *Summa Theologica*, "Habitudo ad causam non intrat in definitionem entis quod est causatum." (The relation to a cause does not enter into the definition of the thing that is caused.) (1a. 44, art. I) and so someone may argue that:

> Nihil prohibet inveniri rem sine eo quod non est de ratione rei: sicut hominem sine albedine. Sed habitudo causati ad causam non videtur esse de ratione entium: quia, sine hac, possunt aliqua entia intelligi. Ergo, sine hac, possunt esse.
>
> (Nothing prevents a thing's being found without what does not belong to its concept, e.g. a man without whiteness; but the relation of caused to cause does not seem to be part of the concept of existent things: for they can be understood without that. Therefore they can be without that.)

This strips the argument (as it occurs in Hume) of the damaging experimental premiss about what we can conceive without also thinking of what. Nor is there here an argument from imaginability to possibility. Rather Aquinas is using the idea of the *ratio* (Greek *logos*), which I have rendered 'concept', of a thing.

Thus the argument published by Aquinas goes:

> It is possible to understand existents without the relation of caused to cause
> ∴ That relation does not belong to the concept of existents
> ∴ They can lack that relation.

Whereas Hume's version goes:

> It is possible to conceive an object coming into existence without conjoining the distinct idea of a cause
> ∴ It is possible to imagine: an object coming into existence without there being a cause
> ∴ It is possible for an object to come into existence without a cause.

There is a great similarity; an important difference is the idea of the grasp of a concept which gives what is essential to a matter, which we find in Aquinas' argument. (And something like this we have already looked for in our attempt to introduce 'proper physical statements' of the occurrence of a cause and its effect.) This enables Aquinas to avoid the more obviously false or inconsequent bits of Hume's argument. It thus appears to convey what is acceptable, what has been found convincing, in the suggestions of the passage from Hume.

However, the thing that in Hume's argument is said – as we may now put it – not to involve a relation to a cause is not the thing that is caused, but its coming into existence. So what we have to consider is whether this premiss:

> The relation to a cause is no part of the concept of the thing that is caused

entails

> A thing can begin to exist without a cause

or again whether we can quite simply justify the assertion:

> The relation to a cause is no part of the concept of a thing's coming into existence.

Hume is so satisfied with his own thinking on this question that he deals cavalierly with the only serious argument that he mentions on the other side. It comes from Hobbes; we should be grateful to Hume for bringing it to our notice. As he renders it:

> All the points of space and time in which we can suppose any object to begin to exist are in themselves equal; and unless there be some cause, which is peculiar to

> one time and to one place, and which by that means determines and fixes the existence, it must remain in eternal suspense; and the object can never begin to be, for want of something to fix its beginning. (Book I, Part III, Section III)

This account is a moderately good rendering of the argument as I have been able to track it down in Hobbes (for Hume gives no reference). It may be found in a treatise on Liberty and Necessity; the precise passage is to be found on page 276 of volume 4 of the *English Works of Thomas Hobbes*. It runs:

> Also the sixth point, that a man cannot imagine something to begin *without a cause*, can no other way be made known, but by trying how he can imagine it: but if he try, he shall find as much reason, if there be no cause of the thing, to conceive it should begin at one time as another, that he hath equal reason to think it should begin at all times, which is impossible, and therefore he must think there is some special cause why it began then, rather than sooner or later; or else that it began never, but was *eternal*.

The respect in which Hume's rendering is a bit unfair is that it suggests the absurd picture of an object's hovering in the wings, as it were, and waiting to get on the stage of existence. But I think we would not want to take any notice of that in any case.

Hobbes's argument is a very interesting one and deserves far more attention than Hume bestows upon it. My interests are not exegetical, so I will not go into the question whether all I see in this argument is really what Hobbes meant. I will only say that consideration of Hobbes's argument led me to the following:

First: in general, the place and time of something's coming into existence are independent of the thing in question. I mean that the place and time in which something comes into existence exist, whether or not that thing, or (in case there is any difference) *such* a thing, does come into existence then and there.

Second: space and time are relative. It is nonsense to think of totally empty space – as Einstein once characterized it, an empty sideless box – consisting perhaps of, or as it were stakeable out as, a set of positions. And it is equal nonsense to think of time as another sort of space (a one-dimensional space) that can be conceived to be empty of existent things and also has its fixed positions for things or events to occupy. Equal nonsense, too, to think of the real existence of a space conflated of space and time together, space-time, as an empty space with such a set of positions. It is also nonsense to think of time as a flowing somewhat, as is suggested to the imagination by the phrase "the passage of time".

Times are made by changes in things and in their relations, places by the configurations of extended objects and their parts.

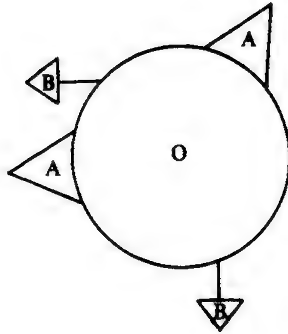
From these considerations it follows that a thing's coming into existence at a time and in a place requires at least the existence of other things. That is to say that it cannot be supposed without supposing the existence of other things as well. This, however, is not yet to give what Hobbes's argument

seeks to give; the necessity of a cause. For the demand that whatever has come into existence has come into existence at a time is satisfied if we can say: it came into existence a time ago. And similarly for future coming into existence: that will be a time hence. And if, without thinking of past or future, we think abstractly of a thing's coming into existence, and think "It has to happen, if it happens, at some time", what we are thinking is that it must be a datable event – 'in principle', as they say. And that demands that there be other things, but not yet a cause.

Similarly for place. But there is not the same universal necessity for a place, as for a time, of a thing's coming into existence. There would be a difference between the two for the first thing that came into existence. We can at any rate construct that description and reflect on its implications. Grant a first beginning, without an eternal container: "place" would not apply to it. We might think that the thing's place is no different from its time; we have the time at least as 'for some *n*, *n* periods ago', even though there could be no prior or contemporary event to give its birth a time. But 'from here', the spatial analogue of 'ago', does not yield much fruit. For how could that – or anything else – be taken as a fixed point of spatial reference by the help of which to assign a meaning to place designations at a time at which nothing fixing that point yet existed? 'Now' of course is not a fixed point of reference either. Nevertheless, in relation to the immense periods of astronomy it is as good as that – one feels no need, for example, to adjust a length of time ago given as 3×10^{16} secs because the book it was given in appeared some years back! But in any case we would know if we needed to how to add periods to our time of the world as everything gets older. Perhaps, *if* it were on the cards to say how long the universe has existed, the answer would be in some way relativistic (I don't know if this might be so); but there is an *a priori* difficulty of a different order about assigning a place to its beginning, i.e. about the idea of any place at all.

An extended object *O* of any kind coming into existence and being the first thing to do so would make at least two places: an indeterminate one, only specifiable as "outside" *O*, and bounded by the surface or limit of *O*, and a determinate one inside *O*, also bounded by its surface or limit. If now our object was a uniform sphere, those would be the only two places it would make. But, in supposing and representing such a thing we can construct various geometrically characterized differentiations both of the space outside and of the space within. These are possible shapes, partially bounded by the surface or a segment of it, or by lines drawn through points specified in terms of *O*. But no point on the surface of *O* will be differentiable from any other and so, unless something of such a shape and relation to the sphere actually comes to be, there will be no difference between potential places defined by congruent shapes and similar relations to the sphere. No difference, I mean, between the places *A*, or again between *B* and *B*, which are severally *represented* as different in the drawing on page 156. These two distinguished places (namely *A* and *B*) would be purely imaginary

because *outside O* is indeterminate. For places within the sphere, like constructions would give similar distinguishable sets of indistinguishable places, members of each of which might be illusorily distinguished in a diagram. But the places within (whether distinguishable or not) are not imaginary because there is a determinate whole to suppose divided into parts.



All this merely to illustrate the principle: no object (or happening), no place. So an object, coming into existence and being the first, or an event which was the first, could not be thought of as coming into existence or happening at a place; it would rather itself bring about the existence of a place. It is the same if it is the universe, the whole world, we are talking about, and we suppose it to have come into existence. Supposing it to have had a beginning means supposing it to have come into existence not out of anything. For if out of anything, that would have already existed and so this event would after all not have been the beginning. I take the universe to be the totality of bodies and physical processes, together with whatever is contained in any manner in the compass of that whole. And so if we take any starting-point by referring to specifiable bodies and processes, the universe will include anything that such things have come out of by way of any sort of development. Thus if the universe had a beginning, it must have come to be out of nothing, i.e. not out of anything. And if we ask "When and where?" we have to make a difference between the two questions. "Where?", as we have seen, can have no sense given to it. And so, if it is the whole universe we are talking about, we have to depart from the principle that whatever comes into existence must come into existence somewhere. But to "When?" the answer will be "For some *n*, *n* periods ago of whatever process we may use to measure the times of the world".[1] Nothing will then have taken place more than *n* periods ago.

[1] I will continue to use this form for the sake of simplicity. Imagine a prisoner in a dungeon who keeps a tally of certain recurrent events *E*, but these cease, and he starts another one, of some other recurrent events E^1. Then he might say that he last saw someone 160 E^1's plus 370 *E*'s ago. I don't know whether anything analogous might occur in the matter of the age of, say, the galaxy.

Thus in this one case we reject the demand that there be a place. There are no places if there aren't any extended things (including processes): their mutual spatial arrangement makes places. There is no such thing as space (the "empty sideless box" that Einstein spoke of) *in* which they occur; but the supposition of their occurrence makes us represent an imaginary indeterminate place around them, and that imaginary place is what we call "space".

However, we still accept the demand that there be a *time when*, because the form for giving it is "so and so long ago". And so we see that Hobbes's argument does at least suggest my present one. A beginning of existence implies the existence of something other than what begins to exist, even if the implication is satisfied merely by processes within that. If, indeed, we are speaking of the beginning of existence of the universe, that is the only way the implication can be satisfied. To repeat, my supposition of the uniform sphere which was the first thing that began to exist was purely for the sake of certain considerations about place and space. I was not suggesting that we could imagine *this*: a uniform sphere is the first and only thing to come into existence, and it suffers no development. If someone says he can imagine this, I will consent to him and say I can too: but such imaginings signify nothing in the way of possibility: they are rather bits of symbolism. The point of my uniform sphere could indeed be put otherwise. Let us suppose a uniform sphere. What places can be specified in terms of it? The answer is: inside and outside, and by geometrical construction as I described. Like figures with like relations to it will not determine different places except in terms of something else that exists in spatial relation to it.

All this, however, is not yet to reach Hobbes's conclusion. For the 'other things' involved in a beginning's having a time may be subsequent to it and so not causes of it as Hobbes intended the word "cause".

Before turning to this, I will dwell for a little on the temporal expressions which may be used here. "Before the world began" is a temporal expression which we use if we argue, e.g., "There cannot have been any things or processes before the world began, out of which it developed". So it might be said "before the world began" would have to designate a time. Again, if the world had a beginning, let it have been *n* periods ago, i.e. let there have been *n* and only *n* cycles or periods of some periodic process which we are assuming as our clock for the universe. We said "Nothing will have taken place more than *n* periods ago". But will not "more than *n* periods ago" also designate a time? And similarly for "$n+1$ periods ago": for we may say "Nothing happened $n+1$ periods ago".

Also someone may have the following difficulty. If there was a beginning of all things then "First there was nothing, then there was something", as the lady said in Disraeli's novel. But that requires that "There is not anything" has been true *then*. And how could that be unless there was a 'then', a time? This is not the difficulty sometimes raised, how a proposition (a sentence) could have been true without existing. For there is no difficulty about that: there are many predicates which do not require the existence of their subject

at the time when they hold of it. But must we not here be referring to *a* time–a supposedly "void time", as Kant called it?

I think that these difficulties about 'then' are all spurious. "A time" should mean a unit period or a number of unit periods, a long time or a short time; or else a point of time. But "before the beginning" or (on the supposition of n periods) "from $n+1$ to n periods ago" cannot designate any length of time, or "$n+1$ periods ago" a point. All that is said is, e.g., that before there were processes there were no processes, i.e. that *this* is not the case: there were processes before there were processes. We might use the contrast between internal and external negation, and say: We are not saying: "At t (t being a time designated by 'before there were processes') there were no processes", but rather "Not this: at t there were processes". Given that n periods is the age of the universe, then "happened $n+1$ periods ago" is not senseless, however, because of the conceptual possibility of the world's having existed for more than n periods.

If there is a real residual difficulty, it will arise only from the idea of a truth which existed then, if we want to speak of this as opposed to only having a past-tense truth which exists now. We must also assume that a mere truth cannot exist alone: for otherwise there would be no difficulty. For the time before the world, would be the time of the existence of the truth that there was no world. However the idea of the existence of a truth and nothing else is unacceptable; because truth is because of the way things are. Note that this is not after all a difficulty about *then*; if we could understand the existence of that truth, the 'then' would offer no difficulty. The truth that there was no world would not be an object or process out of which the world developed, so if it could be supposed to exist it would have existed in a quite different manner from the objects or processes of the world. There would be no length of its duration, nor any temporal differentiation within it. But we cannot accept the mere existence of a truth: that is the difficulty. It was worth considering what would hold if we could, for we see that we would have introduced something whose temporality was altogether different from that of the objects and processes in the universe.

If, then, this is a real difficulty, it can only be solved by postulating some *other* existence or existences which were there before the objects and processes of the universe began, but were not any sort of objects that turned into the bodies and processes of the universe. Remember that I did not explain the "universe" as "the totality of all things", as is often done in spite of the dubiousness of such an explanation. I explained it as "the totality of bodies and physical processes, together with whatever is contained in any manner within the compass of that whole". This makes it possible to postulate other existences, if there is reason to do so. To describe them as existing before the beginning of the world would not *eo ipso* be to ascribe any temporality to them. For *that*, a mutability in them would also have to be supposed; for without change, duration can mean nothing besides existence itself. Except that, if there are processes going on, and there is an unchanging object which

can somehow be compared with those processes, so as to be found or thought of as simultaneous with them, a fictitious distinct idea of its duration arises, as if one could distinguish temporal parts within the 'invariable' object.[2] Now whether anything can be said about such postulated existences, which would be non-temporal or else whose temporalities would at any rate be different from that of our 'clock of the universe', I do not know. And I am uncertain whether the particular difficulty about the existence of truth before the beginning (if there was a beginning) *is* a real or a spurious difficulty. I am only sure that *if* it is a real difficulty, it can be solved only by such a postulate.

An expression that is sometimes on some people's lips is "before time began". That might seem quite absurd, and in two different ways. First, it sounds as if time were being thought of as that 'equably flowing' thing that Newton spoke of: itself a process. Second, as "before" is a temporal expression, how can we speak of anything as 'before time?' At best it must be a violent and contradictory metaphor, like "outside space". But there is another understanding of it which neither involves thinking of time as a process nor invokes a violent metaphor. It may be understood in the sense 'before *times* began'. A time is here to be understood as a period, a 'length of time', an age, or again as a point. For there to be *times* there must be processes measurable by some master time-keeping process. A point of time is a derivative notion, for which it is assumed that the specification of a length of time between two events, or back or forward from now, can be precise.

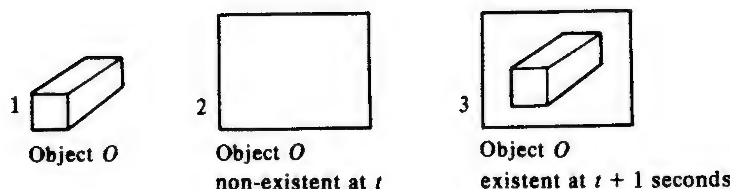
I will now leave these questions, raised by the conception of a beginning of the world, where it is indeed very difficult not to flounder and flail about, gasping for breath and uncertain of talking sense. In all other cases we can accept Hobbes's assumption that when we conceive of something coming into existence we conceive it as coming into existence in some place and at some time. This certainly involves the existence of other things and so falsifies Hume's contention "That there is nothing in any object consider'd in itself, which can afford us a reason for drawing a conclusion beyond it" (Book I, Part III, Section XII). Unless indeed 'something's coming into existence' is not to be reckoned an 'object' as "object" is meant in that dictum – but we have seen that a coming into existence *is* called an "object" in Hume's usage.

As we have so far understood the existence of other things to be involved in something's coming into existence, it does not yet imply the existence of a cause. To this question, then, we must now address ourselves.

We are to try conceiving the beginning of something without a cause. It is clear that Hobbes meant really supposing this to happen, not just forming a picture of it as happening. In discussing Hobbes's argument Hume rather curiously writes as if what were in question were a proposed existence: "The first question that occurs on this subject is always, *whether* the object shall exist or not: The next, *when* and *where* it shall begin to exist" (Book I, Part III,

[2] Here I am following Hume.

Section XII). Presumably he meant an existence that one proposes to one's imagination. But what one ought to propose to one's imagination is perhaps not the existence of some object, but oneself seriously judging an object to have come into existence. And here again Hume is over-easy with his argument from imagination. As we have seen, he says "'twill be easy for us to conceive an object to be non-existent one moment, and existent the next, without conjoining to it the distinct idea of a cause" (Book I, Part III, Section III). It is very easy to accept this. Till recently I have done so, I have thought nothing easier than to imagine an object non-existent one moment and existent the next. One can make a picture of it:



and of course one can do this without thinking of a cause. I criticized the suggestion that being able to imagine something-coming-into-existence-without-a-cause proved anything, or was proper material for an argument from imagination. I said that "something coming into existence without a cause" was a mere title one gave to one's mental picture of something – a rabbit, say, or a star – coming into existence. But I did not then notice that just the same is true of the description "something coming into existence".

Following Hobbes, I am to try and imagine – really imagine, i.e. imagine the serious supposition, that some object has come into existence without any cause. Now what reason have I, on this supposition, to assign one time and place to this coming into existence rather than another? Can I just suppose some particular time and place without more ado? Not if what I am to propose to my imagination is that I am truly judging that this object came into existence. I need to envisage myself as having reason to say it came into existence at this time and place and not at any other.

Note that we should not raise this question merely about the object's *being* at a certain place at a certain time. *That* I could have observed to be the case; and if I did, I certainly should not need to make reference to any cause of its being at that place then, in justification of my fixing the object as being here now, there then, and so on. But I have got to suppose the object to have been not existent anywhere one moment, existent in this place the next. How can I do that without supposing a cause which justifies me in judging that *that* was the time and place?

"Well," says the counsel for the prosecution (of Hobbes): "Can't you *find* that the object was not in that place at one moment and that at the next moment it was there? Might you not learn that this was so from those who

observed it? Or, better, may you not suppose yourself to make the judgement because you observed this object coming to be at that place yourself?"

But we must notice that there is a difference between coming to be at a certain place – that is, its coming about that a thing is at a certain place – and a thing's coming into existence at that place. So the question arises: granted that it has come about that this thing is at this place (where it was not before) why is this a beginning of its existence? Might it not have existed previously elsewhere and arrived here now?

"Of course it might," the prosecution agrees with some impatience. "But it would have to have travelled, i.e. traversed a path and arrived at this place by approaching it. You could have been observant enough to exclude that."

But might it not have arrived from elsewhere without traversing a path from there to here, simply by being first there, then here?

"No! That doesn't happen. At least, we are told strange things about α-particles – but we aren't dealing with them here, but with familiar macroscopic objects. And such an object could not be identified with such an object – even of just the same kind – that had been in another place, unless there was a spatio-temporal continuity between them."

But I don't have to make such an identification in order to justify my doubt. Making such an identification would of course prove the doubt true, and I would no longer *doubt*. But I am on the contrary trying to justify myself in saying that this object came into existence here and did not arrive from elsewhere. So in such a case my task, it seems, will always be to make sure that *any* identification of this as this individual with something that was somewhere else before is excluded. I have got to be assured both that such objects never behave as we are told α-particles do and that this one did not travel from anywhere else in some other form – say as a gas – and resume the form it has here on arrival. The task is too much for me; and for all I have a right to judge in the matter, I am forced to conclude that this object may have come into existence in any place and at almost anytime you care to mention. It seems that there is no experience which itself *positively* indicates that I have to do with a beginning of existence here, except indeed that the object is here now and was not here before. But, that being obviously insufficient, I have got to exclude other explanations of its arrival here. That it came into existence here is apparently to be arrived at by elimination. Or is there any more direct method of judging the question?

The counsel for the prosecution may well be excused if he is goaded by this question into a prolonged exposition and expostulation: "*Of course*", he says "there is over and over again positive knowledge which shows you that this is a first beginning of the existence of an object – it's hardly ever a question of elimination at all. Take a chair, say, of course it doesn't arrive like an α-particle or in any other strange fashion, but you know that because you know how it does come to be; you might have seen the joiner making it out of the wood it was made of, and the wood sawn into planks and the trees cut down, from which it came. Take a baby, it comes from the parents, it grows

out of the conjoined sperm and egg. And metals are smelted out of rocks and moulded into pots and rings and other ornaments. Consider the flowers: you yourself took a cutting from a parent plant and planted this one here, and you also started this mustard and cress patch with seeds on a bit of flannel, and you watched it grow. Even without going down a mine we know that the separate lumps of coal were hewed out of the coalface; and we ask a geologist about the process of formation of the coal in the ground. The pudding we are about to eat did not suddenly arrive complete, we know its origin, for it was made in the kitchen out of its ingredients by your sister, you watched her do it. And you know equally well that glass was manufactured out of its ingredients and the clothes you wear were made in a factory out of cloth that was woven out of threads that were spun out of fibres that were, for example, gathered from the cotton plant or teased from the fleeces which had been sheared from sheep. So in a hundred cases we know we can observe beginnings of new items because we know how they were produced and out of what. It is preposterous to claim that no positive actual or possible experience reveals beginnings of existence. We know the times and places of their beginnings without cavil because we understand their origins."

The defence need make no reply.

The defence rests with the final speech for the prosecution.

15 Soft Determinism

Keith Lehrer[1] has the following argument. Consider the three propositions:

A (1) If causal condition C obtains, A will φ
(2) Unless causal condition C obtains, A can't φ
(3) C does not obtain.

These three are consistent. But from (2) and (3) there follows "A can't φ". (1) therefore cannot be the analysis of "A can φ" – no matter what condition "C" is a dummy for. Therefore in particular "A can – i.e. is free to – φ" cannot be explained as "If A chooses to φ, A will φ" where choice is understood as a causal condition. (This is the standard understanding in the analysis of freedom of will proposed by Moore.)

Lehrer's argument has been criticized by Alvin Goldman[2] on the grounds that it could be used equally well to fault the analysis of "X is soluble in water" as "X will be dissolved if it is immersed in water". "Let X be a piece of sugar, and imagine a magician such that if X is not immersed in water he changes its molecular structure so as to make it non-soluble, whereas if X is immersed in water it dissolves." On this supposition it would seem that three consistent propositions parallel to the A-triad could be framed, and Lehrer's argument used to fault the conditional analysis of "X is water-soluble". But the conditional analysis is certainly right, so Lehrer's argument must be wrong!

Donald Davidson also says: If Lehrer's

> reasoning is sound, it shows that no attribution of a power or disposition is ever equivalent to a conditional (whether the conditional is construed as causal law or subjunctive). Thus on Lehrer's argument to say something is water-soluble can't mean it dissolves if placed in water, since it may now be soluble and yet placing it in water might make it insoluble. There are various ways one may try to cope with this point. But for my purposes it will be enough to remark that even if Lehrer's argument showed that no disposition or power is correctly analysed by a subjunctive conditional, the claim that being able to do something is to have a causal power would not be undermined.[3]

Davidson believes that freedom to act is a "causal power". A causal power is "a property of an object such that a change of a certain sort in the object causes an event of another sort".

[1] Keith Lehrer (ed.), *Freedom and Determinism* (New York, 1966) p. 196.
[2] *A Theory of Human Action* (Englewood Cliffs, NJ, 1970). p. 199n.
[3] Donald Davidson. "Freedom to Act" in *Essays on Freedom of Action*, ed. Ted Honderich (London, 1973). Davidson disclaims dependence on the hypothetical analysis of freedom.

From G. Ryle (ed.), *Contemporary Aspects of Philosophy* (London, 1977).

Clearly Lehrer's argument disturbs. Let us examine Goldman's application of it to water-solubility. The three propositions

B (1) If C obtains, A will dissolve
(2) Unless C obtains, A is insoluble
(3) C doesn't obtain

are consistent only if "A is insoluble" is given a sense other than that of the contradictory of (1). Goldman implicitly does this, and Davidson's remarks too imply that it can be done. If we say that, e.g., A's having a certain molecular structure is sufficient for A's insolubility, even though (1) is true, then that is already to abandon (1) as the analysis of "A is soluble".

Someone who held by the hypothetical analysis *should* say that the B-triad is inconsistent. We would then have two opposing arguments, one running "The B-triad is obviously consistent, therefore the hypothetical analysis is wrong" and the other: "The analysis is obviously right, therefore the B-triad is inconsistent."

There is, I think, no indication in Lehrer's article that he would want to apply his argument so as to fault the hypothetical analysis of solubility and similar properties. Nor is this to incur Schopenhauer's stricture on philosophers who treat an argument like a cab – take it as far as you want to go, and then pay it off. It is not so clear as Goldman and Davidson think, that if Lehrer's argument is valid it applies to the analysis of dispositional properties. For there is a quite different application of it to the case of water-solubility, which is the real parallel, and is in various other ways more relevant in the original argument. Consider:

C (1) If C obtains, A will dissolve (in water)
(2) Unless C obtains, A can't dissolve (in water)
(3) C doesn't obtain

When we put Lehrer's original argument side by side with this one, it hits us on the nose that what he has to defend against is the charge of a fallacy of ambiguity. We have, we will suppose:

A (1) If C obtains, A will φ
(2) Unless C obtains, A can't φ
(3) C does not obtain.

All right, A can't φ, but how do we know whether *this* is – or entails – the negation of the "A can φ" which expresses A's freedom of will? By stipulation? Lehrer knows that all soft determinists hold that in one sense (the free-will sense) one can act otherwise than one does, while in another sense (deterministic causal necessity) one can't. In the case of the dissolving of A, we see quite clearly that the possibility expressed by "solubility" is not what is in question in C (2): "Unless C obtains, A can't dissolve".

"If A is immersed in water it will dissolve" cannot really be adequate to the intentions of those who offer it as an analysis of "A is water soluble". They certainly mean it will dissolve in the water it is immersed in. Hence we have to quantify. I will take: "x is W" to mean "x is ordinary water under ordinary conditions", and I will take "immersion" to cover, say, being soused, being sprinkled, etc. Then, instead of "If A is immersed, it will dissolve", we put:

D (1) For all x, for all t, if x is W then if A is immersed in x at t and left in x, A will start dissolving in x within t + i,

where i is some small time interval. And instead of "Unless C obtains, A can't dissolve in water" we put:

D (2) For all t, *if* there isn't an x which is W at t and in which A is immersed and left at t, *then* for all x which is W at t it is not possible that A starts to dissolve in x within t + i,

and instead of "C does not obtain" we put:

D (3) There isn't an x such that x is W and A is being immersed in x just now.

From all of which we may infer:

It is not possible that A starts to dissolve in water now or within now + i.

We may call this impossibility the lack of a possibility for A, and hence, if we like, the lack of a power, or a sort of impotence, on A's part. But it is not the lack of disposition, a general power or capacity.

There is nothing objectionable about the D-triad or the propositions of our C-triad, explicated in some fashion; certainly they are not incompatible. And there is nothing to object to about the conclusion.

Note that in the C-triad we have something illustrative of the plausibility of the general principle, that a causal condition of something may be a necessary causal condition of its possibility. Indeed, in the D-triad it is not just a bare possibility that the condition mentioned should be a necessary condition of the possibility of the result. It is *necessarily* the case that it *is* such a necessary condition, because of what dissolving is. We could however construct cases where this was not so. For example, we could suppose, if we wanted to, that someone would die if he took prussic acid, and that that was the only way he could die. Then, given that there was no prussic acid around for the time being, then just for the time being he couldn't die.

Note also that, once we have granted the consistency of the C-triad, we cannot offer C (1) as an analysis of the "A can dissolve" whose contradictory is the consequent of C (2).

Back now to the original argument. When "A can φ" expresses A's freedom, which is it analogous to, "A is soluble" or "A can dissolve"? Evidently to the latter, in the following way: "A is soluble" may still be true where there is obviously no possibility for A of actually dissolving. But "A can walk" (where this expresses A's freedom in respect of walking) is certainly not true where there is obviously no possibility of walking for A because, e.g., he is tied up.

It might be thought that "A has free will" is analogous to "A is soluble".

But this cannot be right. "A is soluble" has a certain relationship to "A (here and now) can get dissolved". What stands in the same relationship to "A (here and now) can φ" would be a proposition like, say "A has a general capacity to φ". For example, let "φ" = "walk". Then "A can walk" may express A's competence to walk; he knows how, he is not a cripple, he has not a broken leg. Freedom of will, or even freedom of will in respect of walking, is thus not the analogue of water solubility. The analogue to that for walking is rather a general capacity for which we haven't got a single term in English; but let us coin one and call it "ambulability". There is no such thing as a general capacity, which we would describe as "freedom of will in respect of walking" which like "ambulability" may be still attributable to someone who can't walk because he is tied up. If he is tied up, he has no 'freedom of will' to walk.

We are now in a better position to assess the argument. The suggestion being considered is that choosing is a 'causal condition' of any action that is called free. "He can, he is free to, act so – I mean, in the sense of doing it if he chooses" one might say, and this has been taken as an analysis as a conditional, and the analysis interpreted by Lehrer and many others as a statement of a *causal* condition. We now consider the three propositions which, according to Lehrer, must be compatible if C is such a condition:

E (1) If C obtains (i.e. if A chooses to walk now), A will walk now
(2) Unless C obtains, A cannot walk now
(3) C does not obtain.

Here it is clear that the "A cannot walk" of E (2) does not refer to A's lacking the general capacity to walk. What E (2) says A can't do, unless C obtains, is *walk now*: the proposition is not to the effect that unless C obtains, A now lacks the general capacity.

The neo-Mooreian doctrine considered by Lehrer amounts to this (in the particular case). A can walk: this is supposed to express his freedom of will in respect of walking. Then A has, not only a capacity of walking, but a present possibility of walking; he is not tied up, for example. Now there is of course a further set of causal conditions, which includes various things going on in his muscles and nerve fibres, such that, *if* they are all actualized, A will walk. Among these is supposed to be included a choice on A's part, a choice to walk.

Lehrer's argument is that, for any such condition, it is logically possible that (like not being tied up) it is a necessary causal condition of A's here-and-now ability to walk. The statement that it *is* such a necessary condition of possibility is perfectly compatible with the statement that it is not actualized, and together these would imply that it is not possible for A to walk, that A (here and now) cannot walk.

And now the question is whether this "cannot" contradicts the "can" in the "A can walk" that expresses A's freedom of will in respect of walking. Lehrer's opinion is apparently that it does. That is, that the absence of *any* necessary condition of the possibility of exercising a capacity of walking (say) is sufficient to prove that one is not free to walk. Just as not being tightly chained up is a necessary condition of the possibility of exercising one's capacity to walk, and so in the absence of that condition – i.e. given that one is tightly chained – one does not possess freedom of will, to walk.

The picture of choice as a causal condition of doing something – a causal condition, given which one will do it – is a picture of it as like the last added weight which will start moving the weight on the other side of a pulley. We must emphasize that Lehrer is not arguing that choice *will* then be a necessary causal condition of the event or of its possibility, any more than that the addition of this weight is a necessary condition of the motion of that one. He is arguing merely that its being a causal condition at all is logically compatible with its being a *necessary* causal condition of the *possibility* of the event. And that, since that is so, its being a causal condition at all is compatible with propositions whose truth would entail a statement of impossibility which is incompatible with the possibility expressed by "A is free to φ".

Everyone will allow that "A can walk, i.e. has freedom of the will in respect of walking" would be gainsaid by A's being chained up. If choice is a causal condition *it is then logically possible that the absence of this condition should be just as destructive of freedom to walk as being chained up is*. Assume this logical possibility to be actual, and assume that A does not choose to walk. It will follow that he is no more free to walk than if he were tied up. Therefore – so Lehrer's argument goes – "If he chooses, he will walk", which may be true compatibly with those assumptions, cannot be the right analysis of "He is free to walk".

Now can we say: with these considerations, the charge of a fallacy of ambiguity is rebutted? If being chained up is incompatible with freedom of the will (in respect of walking), then on the assumption that choice is a causal condition, so equally *may* absence of choosing be.

May it? *Why* is being chained up incompatible with freedom of the will in respect of walking – that is, with freedom to walk? Is it because absence of chains is a causally necessary condition of its being possible (here and now) to walk? – If so, won't the absence of *any* causally necessary condition of the possibility of an action be incompatible with freedom to act?

Ah, but that is just what the soft determinist denies. Of course he denies it! There are many causally necessary conditions of the possibility of walking, and in the absence of any of these, he grants, one cannot walk. Some of them are external, and in the absence of these, he says, one not merely cannot walk, but this "cannot" contradicts freedom. This holds of some internal conditions too; one must not be paralysed, for example. But it does not hold of all internal conditions. There are some in the absence of which – of course – one cannot, but still in the *freedom* sense of "can", one *can*. Lehrer has merely pointed out abstractly the possibility that choice should be, not merely a condition of a man's φ-ing, in that sometimes if a man chooses to φ, he will φ, but also a necessary condition of his φ-ing's being possible. If the

conditions are causal and if this abstract possibility is actualized, then, for a man who does not choose to φ, φ-ing *is* causally impossible. "He cannot φ" will be true for as long as he does not choose to φ. But this, the causal-deterministic sense of "cannot", is *not*, so the soft determinist says, the negation of the free will sense of "can".

And what *is* that sense of "can"? Well, the neo-Mooreian tells us, and Lehrer's argument against his definition and the compatibility of determinism and free will has no more force than use of the C-triad to refute the hypothetical analysis of solubility. He said the A-triad was consistent, and he inferred that that analysis was wrong. But that assumes that the "cannot" of the A-triad negates the "can" that the neo-Mooreian is analysing.

Lehrer's opponent, then, says, or ought to say, that the "cannot" that was being analysed will not fit into such a triad of possible and mutually consistent propositions. Plug the neo-Mooreian analysis of "A can φ" (in the free will sense) into A (2).[4] You get an inconsistent triad:

F (1) If A chooses to φ, A will φ
(2) If A doesn't choose to φ, then not:
(If A chooses to φ A will φ)
(3) A does not choose to φ.

Now if this is so, we can derive the contradictory of any one of the triad from the other two.[5] And *this* carries the argument a stage further. For if that analysis *is* correct, we have the right to read "If A chooses to φ, A will φ", either just as it stands, or replacing it by "'A is free to φ". And this gives some curious results, of which I will give three. (a) From "If A chooses to φ, he will φ; and if A doesn't choose, he is not free to φ" we can infer that A does choose to φ. Now might not a person's choice to do something be a causal condition of his freedom to do it in the following way: if he does not actually choose to do it, there will be external obstacles in his way, which make it impossible for him to do it? It is agreed on all sides that physical constraint and exterior obstacles impair freedom – that if, for example, you physically can't commit suttee because "they won't let you", you aren't free to do it. Suppose, then, that they will let you be burnt on your husband's funeral pyre, but only if you choose. Your choice then becomes a causally necessary condition of the lack of external constraint. Why not then also of your freedom? (b) In the same circumstances, if we know that A does not choose to be burnt, we are apparently able to infer that it is not true that if she chooses to be burnt, she will be burnt. Or again, (c) if we know that if she chooses she will, but that she does not choose, we can infer that no such bizarre circumstances hold. Now all of these inferences are absurd. Therefore

G (1) If A chooses to φ, he will φ
(2) If A doesn't choose to φ, A is not free to φ
(3) At T, A does not choose to φ

are compatible. But the G-triad is the same as the F-triad, except that F (2) has been replaced by G (2); and these are equivalent, if the analysis is right. If, then, the F-triad is inconsistent and the G-triad is consistent, the analysis cannot be right.

It is easy to construct other examples on the same lines. For example, you are free to go over the cliff edge but don't choose to. Does it follow that there isn't anyone or anything that prevents you from doing it unless you choose to?

Again: A will marry B (will go through a marriage ceremony with B) if he chooses. But if he doesn't choose, the ceremony won't be allowed. There are people who can tell whether he is acting voluntarily or not, and if not they will stop it. So if he doesn't choose, external constraints will render the proceeding impossible. Could one infer from this that A chooses to marry B?

These arguments have a certain oddity about them – to which I will return – and also might seem sophistical in the following way: Consider the first two propositions of the G-triad:

(1) If A chooses to φ, A will φ
(2) If A doesn't choose to φ, A is not free to φ.

They have got to be tied to a particular time. There is no question of our constructing analogues of the propositions of the D-triad, where we generalized over all times. What then are we to say? Can we put

(1a) If A at T chooses to φ, A will φ by the time T + i
(2a) If A doesn't at T choose to φ, A is not free to φ between T and T + i.

This perhaps seems unreasonable because, if we make the interval long enough to include all the time within which A's choice at T may be implemented, it will often be unlikely that a choice a little later than T could not be implemented within T + i.

However, this is after all irrelevant. For the argument turns, not on the *truth* of any three propositions which would be instances of the G-triad, but on the possibility of there being such propositions. So the objection would stand only if there could never be true instances of (1a) and (2a). But that can't be claimed. So the argument does not contain a concealed sophistry which will be resolved when we consider time references.

Now for the oddity of the arguments. There seems to be something very

[4] I owe this useful suggestion to my student David Waterman, who used it to confute Lehrer.

[5] But note that if we interpret our conditionals truth-functionally, (2) by itself is inconsistent with (3). This would generally be regarded as a red herring, as these conditionals are thought to be future indicative analogues of subjunctive conditionals and the received opinion is that the "if, then" of subjunctive conditionals cannot be truth-functional. But I do not think that that has really been shown: see my "Subjunctive Conditionals", chapter 18 of the present volume. Briefly, I there argue that what the connectives in subjunctive conditionals connect are not propositions with truth values, but subjunctive clauses which have none; the question whether the connectives are the truth-functional ones can therefore only be determined by considering whether the usual equivalences hold, and in fact they do hold.

queer about G(2) in conjunction with G(1), and it remains queer when we supply examples. It is strange to think of freedom as impaired or destroyed by such physical constraints as those. We picture obstacles to freedom as obstacles that resist efforts to surmount them or ones that it is not the slightest use trying to surmount – but *those* obstacles were none so long as A *chose* to do what they otherwise prevented.

We made out cases in which choice was a causally necessary condition of the possibility of an event (the burning of the widow, the occurrence of the marriage ceremony, the falling over the cliff). But doesn't it turn out that a causally necessary condition of the absence of external constraint is here *not* the same thing as a causally necessary condition of freedom to do the thing? And yet, as we have observed, external constraint is generally agreed to be incompatible with freedom. How can it be, then, that a causally necessary condition of the absence of constraint is ever *not* a causally necessary condition of freedom?

Our answer to the charge of sophistry removes our difficulty, however. When we speak in a loose and popular way of a man's being free to do something, we probably have examples in mind where, as we might say, he's free for quite a little time ahead. If A doesn't choose to φ at a given moment, it will then be silly to deny his 'freedom to φ' – because, after all, he may choose to do it and do it at any time within some indeterminate span we vaguely have in mind, and that's what we meant in thinking of his freedom. But if the neo-Moorean analysis is the right account, we have got to be a bit more exact in our thinking.

Instead of the F-triad we must put:

H (1) If at T A chooses to φ, A will φ by the time T + i
(2) If at T A does not choose to φ, then not (1)
(3) At T, A does not choose to φ,

which is plainly inconsistent; and instead of the G-triad:

I (1) If at T A chooses to φ, A will φ by the time T + i
(2) If at T A does not choose to φ, A is not free to φ between T and T + i
(3) At T, A does not choose to φ

and then our argument can proceed as before. If the I-triad is inconsistent, it appears that in a particular case we could make deductions as to matters of fact from any pair of the triad, which we plainly should have no right to make. Therefore the I-triad is not inconsistent. Therefore I(2) is not equivalent to H(2). Therefore the neo-Moorean analysis is wrong.

This argument could be resisted if it could be argued that propositions (1) and (2) of the relevant triads were already inconsistent with one another. The conjunction of them is certainly rather difficult to understand in the F and H triads, when the antecedent is taken as stating a causal condition, though I don't know how they could be shown to be inconsistent. A sworn addict of neo-Mooreian analysis might say that the difficulty of understanding the conjunctions shows there is an inconsistency, and then say that this must be transferred to the corresponding conjunctions from the G and I triads. But we have actually shown and explained how those conjunctions might be true, on the assumption that choice is a causal condition. The negation of freedom is guaranteed by *certain* constraints (this is agreed by everyone), and choice could obviously be a causal condition of such constraints.

Let us now take a real case of a condition of voluntary movement. Certain patterns of brain activity are found to occur very shortly before the initiation of a movement in response to an order to perform it (at once). When the performance is 'self-paced' – i.e. the subject chooses when to make the movement – there is also such a pattern; I understand that it is more diffuse and there is some small difference of time interval from what you get in the other case.

Taking these facts, it may be reasonable to suppose (a) that unless there is such a pattern of activity in someone's brain there will be no initiation of movement on his part, and (b) that normally (i.e. if it is not prevented) when there is such activity, there will be initiation of movement. If (a) were very solidly confirmed, we'd say that without this activity the subject cannot make the movement. In short we shall have a good example for Lehrer's pattern of argument. We shall have as a consistent triad:

(1) If there is a Z-type pattern of activity in A's brain, there will be initiation of movement on A's part
(2) Unless there is a Z-type pattern of activity in A's brain, there can't be initiation of movement on A's part
(3) There is no Z-type pattern of activity in A's brain.

The first two propositions, we may say, are probably true, and presumably the third one is often true. Now won't it be true to say that when all this holds, A can't move? And if that is true, doesn't it follow that A is not free to move?

To answer these questions we have to introduce time references, and here we can quantify universally over times. For the sake of the argument, let us assume a constant interval between the onset of the brain activity and the initiation of movement, and let us call this interval i. Then, taking C as the presence of the relevant brain activity,

For all t, if C does not obtain, beginning at t, no initiation of movement on A's part is possible before t + i.

Now let us assume that precisely *now* C is not beginning to obtain. (Perhaps I fire a shot or read a clock to fix the time here meant by "now".)

It follows that no initiation of movement on A's part between that *now* and that *now* + i was possible.

And it also follows that A wasn't able to start to make a movement between

that *now* and that *now* + i – for this is merely a rephrasing of the preceding proposition.

That being so, how can anyone say that in similar circumstances A is *free* to start making a movement between a *now* and the same *now* + i? Obviously one cannot.

But when we say that A is now free to move, of course we are not thinking in terms of such tiny intervals. The conditions we have laid down do not preclude A's beginning to make a movement as soon as you like after *now* + i; i is very small, and if A can make a movement after *now* + i + $\frac{i}{100}$, that would be good enough. Our assumed facts tell us that if A *does* make a movement then C will have begun to obtain at *now* + $\frac{i}{100}$.

To sum up: as soon as we get quite clear about internal conditions, without which a movement on A's part is impossible, the contrast between internal conditions that make an action impossible, and external ones, like being tied up, proves to be illusory. It trades on vagueness and ignorance.

People are divided into two opposing camps, for one of which it seems quite obvious that a physical impossibility of walking (say) contradicts freedom to walk, and that it makes no difference whether the physical impossibility arises from an internal or an external state. For the other, it seems equally obvious that, when the internal state is connected with choice, that makes all the difference. A choice is here thought of as an event: it causes (no doubt among other things) another event which is called a "free action". It is possible that, in the absence of the event that would have been a choice had it occurred, one's internal mechanism is such as to render impossible that event which would have been called a free action if caused by a choice. Then the choice was a necessary condition of the possibility of the action, and it was physically impossible that one should (say) walk. But just because it is *choice* that is such a necessary condition, the soft determinist says that this is the sort of impossibility that does not contradict freedom.

But what then is the explanation of "freedom" as it is here spoken of? There is none offered except *via* the neo-Mooreian definition. And that we have exploded.

I personally found Lehrer's argument convincing against any hypothetical analysis of freedom. But then I have never thought that freedom was compatible with physical impossibility. I realized that the soft determinist would be untouched by the argument because, of course, he does think freedom compatible with physical impossibility. Naturally, since, being a determinist, he thinks that everything except what actually happened was always impossible. Lehrer's "cannot" is therefore compatible with *this* "'can' of freedom". Now, having disposed of the supposed explanation of this "can", I am at liberty to say that I believe a "'can' of freedom" which holds in face of physical impossibility is pure nonsense.

16 Causality and Extensionality

There is a use of a particular sort of argument to show that if a context is extensional – in a special sense which I will give – and if it involves the embedding of one proposition in another one, then that context is truth-functional. All that is required in the way of further assumptions (apart from the usual ones) is that logically equivalent sentences can be substituted for one another in the context *salva veritate*.

What is meant here by the context's being extensional is simply that designations of the same thing can be substituted for one another *salva veritate*. It is hardly to be entertained that this condition would hold and the one about logical equivalents not hold, so the further restriction doesn't seem to be more restrictive.

The argument I have in mind was first produced (so far as I know) by Quine in "Three Grades of Modal Involvement".[1] Related arguments – that is, using the same sort of trick but to different conclusions – appear in the version of his "Reference and Modality" in the second edition of *From a Logical Point of View*[2] and his *Word and Object*.[3] The kind of argument has some association with the name of Follesdal; and another version of the actual argument I am interested in here comes in Donald Davidson's article "Causal Relations".[4]

The essential trick is to produce a designation of a class or of an object, or an open sentence, which incorporates an independent proposition as a conjunct. Our argument uses the description of a class[5] – say $\hat{x}(Gx \cdot p)$, assuming G to be a respectable class-forming predicate. Let G be such that $\hat{x}(Gx)$ logically can't be empty. Then $\hat{x}(Gx \cdot p)$ is the same class as $\hat{x}(Gx)$ iff p is true. Also, the proposition saying that $\hat{x}(Gx \cdot p)$ is the same class as $\hat{x}(Gx)$ is logically equivalent to p itself. But also $\hat{x}(Gx \cdot p)$ will be just the same class as $\hat{x}(Gx \cdot q)$ whatever p and q may be, so long as they have the same truth value. For, if they are both true, the class in question will be $\hat{x}(Gx)$ and if they are false it will be the null class.

Then we have an argument to show that if $F(p)$ is a context in which p is embedded and if the context is extensional in the sense mentioned, it must

[1] *Proceedings, XIth International Congress of Philosophy, Brussels 1953*, vol. xiv (Amsterdam, 1954). Reprinted in *The Ways of Paradox* (New York, 1966). The originator of the argument was Alonzo Church, reviewing Carnap's *Meaning* and *Necessity* in the *Journal of Symbolic Logic*.

[2] New York, 1963; TB 566.

[3] Cambridge, Mass., 1960.

[4] *Journal of Philosophy*, 69, 21 (9 November 1967), 691–703.

[5] Or it could use the description of a number: the number of numbers n such that n is an even prime and p. This number will be 1 if p is true and 0 if p is false.

From *Journal of Philosophy*, 66, 6 (1969).

be a truth-functional context. For, as you can replace a proposition with any logically equivalent proposition *salva veritate*, you can replace p in the context by the proposition saying that the class of *x*s such that Gx and p is the same class as the class of *x*s such that Gx. And, as you can substitute designations of the same class for one another, you can replace the designation $\hat{x}(Gx \cdot p)$ by $\hat{x}(Gx \cdot q)$, whatever q may be, so long as it has the same truth value as p. And then you can replace the proposition saying that $\hat{x}(Gx \cdot q)$ is the same class as $\hat{x}(Gx)$ by its logical equivalent q. So you can derive $F(q)$. So the context $F(\quad)$ is truth-functional.

I suppose someone might raise a howl about the artificiality of the procedure. I am not inclined to do so. The class description here constructed, for example, seems to be as determinate as G and p may be.

Donald Davidson uses an argument of this pattern to prove that the 'logical form' of causal statements cannot be that of a connection of propositions. And similarly for temporal propositions. That is, the propositions:

p before q
p after q
p because q

disguise the true logical form of what is being said by their means. For – assuming that q is indicative in sense in all three cases – the truth of each of these demands the truth of both clauses. But the contexts are extensional. At least, they are if p and q are, which is all I need trouble about at this point. Therefore, by the formal argument, they must be truth-functional. But then you ought to be able to switch the clauses *salva veritate*. But you clearly can't. It follows, then, that the form of a connection between propositions is borrowed clothes, disguising the true shape of what lies underneath.

Before going further than this, let's ask, what *is* the relation between extensionality in the sense here given it, and truth-functionality? The argument considered proves that *if* you have a connection of propositions and extensionality reigns, then the connective is truth-functional. We can't say conversely that if you have a truth-functional connective extensionality reigns: that would exclude joining non-extensional contexts truth-functionally. What you can say is that if you connect extensional contexts truth-functionally, extensionality still reigns – an extensional context isn't going to become non-extensional by being connected truth-functionally with another context. Nor is a non-extensional context going to become *more* non-extensional by being embedded in a truth-functional context.

As far as the argument goes, then, it would seem that the argument gave us a demonstration of non-extensionality from non-truth-functionality, and nothing else at all. That is to say, if the provenance of the argument had been simply the discussion whether causal and temporal connectives constructed intensional or extensional contexts, it might have gone on like this:

If a connective is extensional, it is truth-functional (proved).
But these connectives are non-truth-functional, for the reasons given.
∴ These connectives are nonextensional.

It is only the conviction that the connectives must be extensional that leads to the conclusion that they are bogus connectives, i.e. don't truly represent the logical form of the propositions constructed by means of them.

I am inclined to accept the argument and not look for a way out. Its conclusion was not a suprise to me in respect of causal statements, and was so in respect of temporal ones. About the latter I will observe only that the non-extensionality comes out clearly enough when you have time-linked predicates in definite descriptions. You can have

John met Joan after they were married

without provoking anxiety *de interpretatione*, but

The Emperor's baby son met his wife after they were married

– where it is understood that the 'his' relates to the baby son – provokes the question, "Not *while* he was a baby, surely?" (The temporal connective 'after' here relates clauses rather than independent propositions, but that is immaterial). Thus the conclusion of the argument that temporal connectives make contexts non-extensional is not after all outrageous.

Causal statements are usually simply asserted to be extensional – so long as their component clauses are extensional, that is – without argument. Or perhaps a few examples are offered, and extensionality firmly asserted in respect of them, by way of (or in lieu of) argument. Consider

The child died because Joan is Rhesus-negative.
The child died because his mother is Rhesus-negative.
The child died because the tallest girl in town is Rhesus-negative.

or

There is an international crisis because "moi, de Gaulle" made a speech.
There is an international crisis because the President of the French Republic made a speech.
There is an international crisis because the man with the biggest nose in France made a speech.[6]

In response to the third of each set, just as we said "But not while he was a baby, surely?", one also says, "But not *because* she is the tallest girl etc." or "not *because* his is the biggest nose". Now of course those who believe causal statements to be extensional will give an account of the 'greater explanatory force' of the second member of each trio. But the question here is not whether one can defend a thesis through thick and thin (we knew that already), but really whether there was originally any good reason for this

[6] I owe this pleasing example, as well as the thought about temporal connectives, to P. Geach.

thesis at all. Here I am in a bit of a difficulty. For I have no sure insight into the sources of the conviction that causal statements are extensional.

At any rate the comment "not *because* . . . is the _ _ _" offers a dilemma: either you must grant that this ". . . is the _ _ _" is not an identity proposition, or you must grant that *this* "because" context is non-extensional. For clearly the phrase 'the . . .' can't be regarded as *here* replaceable by some other designation of the same.

One idea that seems to be operating (though it has nothing especially to do with causality) is a crude Fregean sort of view of designating. For on that view we don't have to consider the *scope* of a definite description; definite descriptions and ordinary proper names are the same sort of expression, and the innards – the syntactical complexity – of a definite description don't make any difference, don't force us to qualify the comparison to a proper name. For of course I wouldn't want to deny that there was an international crisis because the man with the biggest nose etc., while granting that it did happen because 'moi de Gaulle' did whatever it was. The case cries out for a Russellian kind of treatment; I mean a differentiation between:

> Concerning the man with the biggest nose: there was an international crisis because *he* . . .

and:

> There was an international crisis because the man with the biggest nose . . .

– the latter, since it is differentiated from the former, now being *false* – unless, as we say, "it was *because* his was the biggest nose".

This treatment involves allowing that difference of scope may make a difference even when a definite description is nonvacuous. In a footnote in "Reference and Modality" Quine contrasts Arthur Smullyan with Russell when Smullyan adopts such an idea to solve Quine's well-known problems about "Necessarily the number of the planets is greater than 7". At least he claims that Smullyan was departing from *Principia Mathematica* while, I suppose, ostensibly working within its framework. For he observes that Russell's theory of descriptions involved differences of scope, but adds that change in the scope of a description was indifferent to the truth value of any statement unless the description failed to name. "This indifference was important" he goes on "to the fulfillment, by Russell's theory, of its purpose as an analysis or surrogate of the practical idiom of singular description. On the other hand, Smullyan allows difference of scope to affect truth value even in cases where the description concerned succeeds in naming."[7]

Now Smullyan can't fairly be charged with just playing fast and loose with the system of *Principia*. For Russell says, "But even when E!(ɿx)(*Gx*), the in-

[7] *From a Logical Point of View*, p. 155, fn 9. *Note while in press*: Quine has corrected the mistake in his latest edition. The Smullyan article referred to is "Modality and Description", *Journal of Symbolic Logic*, xiii, 1 (March 1948): 31–7.

completeness of (ɿx)(*Gx*) may be relevant *when we pass outside truth-functions*."[8] He gives a psychological example, as he usually does; the one exception I know is 'it is a strange coincidence that'[9] – and even that he probably regards as a psychological example, too. But the important phrase is that one: "when we pass outside truth-functions".

Quine could indeed quote various passages from Russell in which he says, in the 'blurb' appended to his demonstrations, that "when (ɿx)(*Gx*) exists, the fact that it is an incomplete symbol becomes irrelevant to the truth value of logical propositions in which it occurs" (see *14.18). But his proofs show that he is *not* "passing outside truth functions" (*14.28; see also *14.3). And so every passage in which he says that, given the existence of (ɿx)(*Gx*), truth values are unaffected by scope, we should take him as being characteristically careless in omitting the qualification that he put in on page 87: the incompleteness may be relevant when we pass outside truth functions, even when (ɿx)(*Gx*) exists.[10]

With temporal and causal connectives we do pass outside truth functions; so even within the *Principia* system it is open to us to allow difference of scope to affect truth value, as Smullyan did for modal propositions.

Quine further observes that Smullyan's suggestion involves a fundamental division between proper names and definite descriptions, and Quine admits that, if you accept that division, then examples that show failure of substitutivity must exploit some descriptions rather than just proper names. He goes on to say that this means adopting an "invidious attitude" towards certain ways of specifying something x, e.g., towards "There are just x planets", and favouring others such as

$$x=\sqrt{x}+\sqrt{x}+\sqrt{x}\neq\sqrt{x}$$

as somehow better revealing the "essence" of the object. This is a "reversion to Aristotelian essentialism".

I wouldn't personally regard that as an objection if it were true – but it is surely quite false. For all that is required in the way of invidious attitudes is (1) the assignment of a peculiar role to the proper name, (2) the treatment of some unique descriptions of an object as not *necessarily* satisfied by it. After all the topic was modality! And that treatment is simply the acceptance as true of such a proposition as "necessarily, $9=3^2$".

This point I make not for its own sake, but because it is clarified by the parallel point about causal statements, and showing how this is so may clarify *that* (latter) point. For when we say, "Sure there was a crisis because the

[8] 1st ed., p. 87; 2nd ed., p. 83. My emphasis.

[9] 1st ed., p. 77; 2nd ed., p. 73. This last quote is about classes, not descriptions, but the point is the same.

[10] Russell's meaning for "extensional function of a function" is different from the sense of "extensional context" that we have been using – he is interested in the replacement of *predicates* by other coextensive predicates (1st ed., p. 76; 2nd ed., p. 73), while we are concerned only with the special case of replacements of *designations* by other designations of the same object – we should not let ourselves be confused by this.

man with the biggest nose etc., but not *because* his was the biggest nose" we are adopting an 'invidious attitude' toward the description "man with the biggest nose" and by contrast we probably show favour to the description "President of the French Republic". Not, however, because the latter description more nearly reveals the essence of "moi, de Gaulle"; merely because it seems to be true that what he did will have caused a crisis because (among other things) he is President of the French Republic, and not at all because of his having such a big nose. And similarly

$$=\sqrt{x}+\sqrt{x}+\sqrt{x}\neq\sqrt{x}$$

is favoured as a specification G of something in

$$\text{Necessarily } (\iota x)(Gx) > 7$$

and

is the number of the planets

is disfavoured, merely because giving the one specification seems to result in truth, giving the other falsehood. "Essence" is not in question in the necessity statement any more than it is in the causal statement. It is probably only because *necessary* properties are often associated with essences that Quine thought you must go in for Aristotelian essentialism to sustain Smullyan's suggestion. But no such suggestion lurks in the offing in connection with causal statements. So it would need to be *shown* that necessity is a concept that presupposes essences.

Reverting, however, to the substantive topic of this paper: note that the proof of truth-functionality given by Quine does not just bear upon causal statements in which two propositions are connected, i.e. causal statements of such a form as

p because q

The proof concerns *any* context $F(p)$ in which a proposition is embedded. Thus it concerns

A brought it about that p

– whether "A" is the designation of an event or of a substance, for example, doesn't matter. If this context is extensional, i.e. if designations of the same object in p are intersubstitutable *salva veritate*, then it is truth-functional; i.e. p can be replaced by any proposition of the same truth value. This monstrous consequence shows that we must either take the context as intensional, or, adopting Davidson's way out, say that it too "falsifies the logical form of causal statements".

As I have indicated, I find it harmless to say that causal statements are intensional. But our considerations lead to raising the following question: What is at stake in maintaining or denying that an effect is properly described or presented in a *proposition?* I feel that something is at stake – but I don't

know what it is. Whatever it is, in this issue one side is probably correctly represented by the insistence on the proposition but I suspect (my hunch is) that the other side is the right one, but is *not* correctly represented by objecting to the presentation in a proposition.

17 Before and After

In what follows I shall be considering compound sentences formed with the word "before" only when they are indicative in sense. That is to say, I shall not be considering sentences like "I caught the vase before it fell" or "I arrived before he ate anything" where the second clause would go into the subjunctive in Latin, showing that it is not implied that he did eat anything.

I "p *before* q" *and* "q *after* p" *are not universally equivalent, and so* before *and* after *are not converses*

Nothing is more natural than to suppose that "before" and "after", in their temporal sense, name converse relations. The converse of a relation R is that relation in which b stands to a if a stands in the relation R to b; so if R and R' are converses, then aRb is equivalent to $bR'a$: any inference which can be drawn from the one can also be drawn from the other. Let us now consider:

(1) If the Parthenon was there before the Dome of the Rock was there, and the Dome of the Rock was there before St Peter's was there, then the Parthenon was there before St Peter's was there.

(2) If St Peter's was there after the Dome of the Rock was there, and the Dome of the Rock was there after the Parthenon was there, then St Peter's was there after the Parthenon was there.

So far (1) and (2) seem to come to much the same thing: in each case we have aRb and bRc and a valid consequent aRc; so it looks offhand as if in each case we were dealing with a transitive relation. But now let us consider the proposition cRa in either case. For (1) this runs "St Peter's was there before the Parthenon was there"; for (2) "The Parthenon was there after St Peter's was there". The first is incompatible with the consequent and hence with the antecedent of (1); the second is compatible with the consequent and the antecedent of (2). From this it follows that "before" and "after" do not stand for converse relations.

II *The logical properties of the temporal relations before and after*

I shall argue that "before" is transitive and asymmetrical and that "after" is neither, and that the connection between them is that if p before q then q after p, but not vice versa. That "after" is not asymmetrical appears sufficiently from the example already considered. "The Parthenon was there after St Peter's was there" is compatible with "St Peter's was there after the Parthenon was there". It follows that "after" is not unrestrictedly transitive either, for though the result "The Parthenon was there after the Parthenon was there" might possibly be held unobjectionable, the following piece of reasoning is not: "I was born after the Parthenon was there; the Parthenon was there after I was born; *ergo*, I was born after I was born."

From *Philosophical Review*, 73 (1964).

It now seems unsafe to leave it to be judged intuitively that "before" is asymmetrical and transitive. To show that it is both, I will introduce a connective T, roughly corresponding to the vernacular "and then".

The idea behind this connective is that a change might be expressed by saying "*p and then not p*"; "Theaetetus was standing and then Theaetetus was not standing".

A first shot at a rule for this might be:

In ". . . T . . ." the insertions in the blank spaces are always to be or include as conjuncts a proposition and its negation, one on either side of the connective. If the said proposition or its negation occurs as a conjunct, that conjunction is to be a main connective within the proposition on that side of the connective T. To assert a proposition whose main connective is T is to assert the propositions on either side of T. Thus to assert a proposition formed with this connective always involves that one is asserting both a proposition and its negation; for example, "Theaetetus was standing" and "Theaetetus was not standing". Here "Theaetetus was not standing" is an example of what I have elsewhere called an "internal" rather than an "external" negation: the corresponding external negation is given by "not: Theaetetus was standing" – which rules out the truth of "Theaetetus was standing" in relation to *any* past occasion.

However, I do not want to introduce a nonstandard negation, which would lead to complications. I can avoid these by further specifying T. This is now to be understood as "It was the case that . . . and then it was the case that. . . ". The insertions in the blanks are now to be or include a *present-tense* proposition and its negation; for example, "It was the case that Theaetetus is standing and then it was the case that Theaetetus is not standing". Thus asserting a proposition formed with T will always involve asserting that, as regards some present-tense proposition and its negation, each of them has held good.

In concrete examples, however, in order to avoid cumbersome verbiage, I shall write or say "Theaetetus was sitting and then Theaetetus was not sitting".

Now it is awkward to have such a restrictive formation rule as the one I mentioned. In fact, given that we can interpret T under this restriction, we can show it to be interpretable for any insertions of present-tensed propositions on either side of it, assuming that any truth function of present-tensed propositions is itself a present-tensed proposition. For we can say:

$$pTq \equiv (p.\sim qTq) \vee (pT\sim p.\, q)$$
$$(p \vee q)\, Tr \equiv (pTr) \vee (qTr)$$
$$pT\,(q \vee r) \equiv (pTq) \vee (pTr)$$[1]

I shall use an axiom:

If it has been the case that p and it has been the case that not p, p being present-tensed, then we must have either

$$pT\sim p \text{ or } \sim pTp.$$

I say that we have "p again" or "repetition" if we have "$pT \ldots \sim pT \ldots p$". This brings out another feature of T. It is not a simple binary connective: it is rather like "either ... or ... or ..." with an arbitrary number of arguments, read exclusively, which is not reducible to the binary exclusive "or". The assertion of any longer string "$pTqTr \ldots$" involves the assertion of any shorter string contained in it.

I define "p before q" as "$p.\sim qTq$". In concrete examples, the p and q in the definiendum will for convenience be given in the past tense: they could of course be processed into present-tense forms preceded by "it was the case that", but this would be pointless as I shall in any case be using past-tense propositions to flank "and then" in concrete examples.

Many, but not all, past-tense propositions are susceptible of translation by simply being cast into a present-tense form preceded by "it was the case that". I think it accords with our intuition to hold that, if a past-tense proposition is true, then, always, some present-tense proposition was true. But my own concrete examples will be confined to such as are susceptible of simple translation.

I now argue that when repetition is excluded "before" is asymmetrical and transitive:

(1) "Before" is asymmetrical. For, if not, we have a case: "p before q" is true, and also "q before p", that is to say:

$$\text{A. } p.\sim qTq$$
$$\text{and B. } q.\sim pTp.$$

That is to say, we have, for example, both p and $\sim p$ having been true, p by (A) and $\sim p$ by (B). These can only be combined as (a) $pT\sim p$ or (b) $\sim pTp$. If (a) we have (for example) $pTq.\sim pTp.\sim qTq$; if (b) we have (for example) $q.\sim pTp.\sim qTq$. In either case, and in any other arrangement in which we might combine (A) and (B), we have repetition. Thus apart from repetition, "before" is asymmetrical. If repetition is admitted, to be sure it is not: it was night before it was day, and day before it was night.

(2) "Before" is also transitive. For if not we have a case: "p before q" true, and also "q before r", but not "p before r". That is (avoiding repetition), we have:

[1] Mr P. Geach suggested these equivalences in discussion when I read this paper at Los Angeles.

$$p.\sim qT\, q.\sim rTr$$

If we have this, we have had $p.\sim q$ true; we therefore have had either $p.\sim q.r$ or $p.\sim q.\sim r$ true.

In the first case we have $p.\sim q.\; rTq.\sim rTr$, which gives us repetition.

In the second case we have $p.\sim q.\sim rTq.\sim rTr$. Since here we have $\sim r$ in the first clause and r in the last, these must be combined either as:

$$rTp.\sim q.\sim r$$

which will lead to repetition, since we have $p.\sim q.\sim rTq.\sim rTr$ or as:

$$p.\sim q.\sim rTr$$

But in that case we have

$$p.\sim rTr; \text{ that is, } p \text{ before } r,$$

So, repetition apart, "before" is transitive.

These considerations are not invalidated by, for example, the relativity of simultaneity of distant events. Certainly it is possible that p and q simultaneously, and so forth, from one point of view and not from another; but we are considering the relations of priority and posteriority and simultaneity when they are correctly reasoned on, being judged from the same point of view.

The reason "before" and "after" are not converses is this: p before q only if first p and not q, and then q (this can be said independently of the definition given above); but although q after p is then true, this is not the only case of its truth. For example, "q after p" is also true if first p and q and then q and not p. "I was a Boy Scout after you were one" is true not only if I became a Boy Scout after you did, but also if, having perhaps been a Boy Scout as long as you, I went on after you stopped. Further: I was a Boy Scout after you were one even if, having both started and stopped before you did, all the same I was a Boy Scout for a while after you became one. Thus p before q entails q after p, but not vice versa.

We can of course construct conventions for "before" and "after" so as to make them converses. For example, we might make "after" the converse of our present "before", so that "q after p" is true if and only if first p and not q and then q. Then we should deny that the Parthenon was there after the Dome of the Rock was. Or we might make "before" the converse of our present "after". Then we should say that Edward VII was alive before his mother, Queen Victoria, was.

Neither procedure seems to have much point at present; so we can reasonably wait until there is a purpose in one or the other – I do not exclude the possibility that there might be one – before concerning ourselves with it.

III "p *after* q" *is not ambiguous, but has alternative verifications*

It might be held that all that emerges from these considerations is that, if *q* concerns anything that goes on for a stretch of time, like a temple's being there, or a human being's being alive or being a Boy Scout or living at a certain address, then, though "*p* before *q*" is unambiguous, "*p* after *q*" is ambiguous. For, it might be said, "James was ill after John was" may mean:

(1) James began being ill after John began to be ill.
(2) James began being ill after John stopped being ill.
(3) James was ill after John began to be ill.
(4) James was ill after John stopped being ill.

Of these, if (2) is true, (1) is true: if James began being ill after John stopped, he began after John began, assuming that John did begin. On the same assumption, if (4) is true, (3) is true: if James was ill after John stopped being ill, then he was ill after John began to be ill. But (3) may be true – James may have been ill after John began to be ill – though none or all of the others are: he may or may not have begun after John, and may or may not have been ill after John stopped. Someone might say "James was ill after John was", having one of these situations in mind, and be heard by someone else who thought of another of them, so that the statement generated a misunderstanding. But that does not mean that the statement is ambiguous, any more than (1) "James began being ill after John began to be ill" is ambiguous because someone might say it while having in mind that James began after John began and before he stopped, and someone else might understand only what the words strictly say. We must distinguish between a statement's covering a variety of cases and its being ambiguous. The reason we are tempted not to make this distinction here is that we are likely to be interested in which possibility is the actual one, and are quite likely to make the statement having just one of these possibilities in mind. If someone else makes the statement, and we ask which of the possibilities is the actual one, we may frame our question thus: "What do you mean? Do you mean . . . or . . .?"

"*p* before *q*" is without prejudice as to whether *p* also when *q*. So it too might perhaps be held to be ambiguous, since it is true both when first *p* and not *q* and then *q* and not *p*, and when first *p* and not *q* and then *p* and *q*; but the accusation of ambiguity is less likely to be made because something – namely that first *p* and not *q*, and then *q* – is constant and so contains what is essential to being before. If *that* is true, then certainly *p* before *q*, whatever else may hold. But with "*p* after *q*" we have no such constant formula which holds in all the cases. Hence the inclination to speak of ambiguity.

To give another example: "*A* is *B*'s brother" means "*A* is male and *A* and *B* have a common parent". It might be taken for granted that *A* was supposed to have both parents in common with *B*; or that we might be interested to inquire whether he had one or two parents in common with *B* and, if only one, whether it were father or mother; and that we might express our inquiry in the form "what do you mean, his brother?" But this fact does not tend to show that "*A* is *B*'s brother" is ambiguous. Ambiguity is quite different: a statement is ambiguous, not when it has alternative verifications, but when it has alternative interpretations, like "The fall of the barometer alarmed him". This might mean that the crash of the instrument to the ground alarmed him, or that the sinking of the mercury level in the barometer alarmed him, and no one would suggest that it is a statement explicable as a disjunction because it is verified by more than one situation.

Now the fact that "*p* after *q*" is not ambiguous, but has alternative verifications – different *causae veritatis* – would have no importance, if we could not use "*p* after *q*" in inference without having to reason separately on each of the possible verifications. But we can. For example, we may say "James was at large after Smith was ill; therefore he could have seen him and observed such and such effects of his illness;" "James was in touch with John after John was working on his novel, so the possibility of plagiarism cannot be ruled out". We can draw the conclusions without knowing whether James was also at large before Smith was ill, or also in touch with John before John was working on the novel; and without knowing whether James was at large only after Smith stopped being ill, or in touch with John only after John stopped working on the novel.

IV *Quantification over times rejected as offering an escape from the problem*

Now it might be argued as follows: it is true that "after" is not ambiguous in "*p* after *q*" (*q* concerning something that occupied a deal of time). "*p* after *q*" means: "a time at which *p* was after a time at which *q*". For example you were in Italy after I was in Greece because – whatever else is true about the periods of our visits to those places – you were in Italy in 1955 and I was in Greece in 1954. Repetition apart, "*p* before *q*" means "There was some time at which *p* such that every time at which *q* was after it", or "A time at which *p* was before every time at which *q*". If we compare:

A time at which you were in Italy was after a time at which I was in Greece

and

A time at which I was in Greece was before every time at which you were in Italy

– which two sentences are related as are "*p* after *q*" and "*q* before *p*" – we see clearly enough that they are not equivalent and so that, as is contended, "before" and "after" are not converses in such contexts; but the point is quite unimportant. The analysis shows how statements of these kinds can be clarified by quantification over times: in the analysis, the relation "before" and "after" appear as relations between times. We shall not speak of one time as being after another unless it is completely after that other: if two times partly overlap, like "1955" and "Winter 1955–6" we shall of course

not speak of one as after the other but shall divide them up into nonoverlapping times. Thus "before" and "after" will be true converses, when it is *times* we are speaking of.

The advantage of this treatment is that we do not have to concern ourselves with the difference between things which take time and things which do not. If James was born before John was, then the time at which James was born was before the time at which John was; but we need not bother about the definite article, for, if the time at which James was born was before the time at which John was born, then for some time t, James was born at t, and for any time t' if John was born at t' t was before t'.

Quantification over times is often a useful and clarifying device. But it would be absurd to treat it as capable of giving a fundamental explanation of "before" and "after". We can easily imagine a primitive community which had no clocks or calendars, and to whom expressions like "six o'clock", let alone quantification over times, would be unintelligible. They might nevertheless have a quite coherent use of "before", "after", "when", and so forth, and be able to give quite precise time specifications thereby, like "Meet me fifteen days hence when the shadow of the pine is shortest". Our use of temporal quantification presupposes the coherence of temporal language at this level, and further presupposes that we have timekeepers keeping pretty good time with one another, so that we can "tell the time" in an agreed way. It thus involves empirical facts which go beyond the facts stated in the ordinary "before" and "after" propositions supposedly analysed. Someone might object that we do not need clocks and watches – we could use the sun – but it is again an empirical fact that two people can agree on the position of the sun; and this also brings in something that is not in the propositions to be analysed when they do not themselves refer to the position of the sun.

V *Preliminary justification for my unorthodox conception of a proposition*

It will have been apparent that in Section II I departed without apology from the most generally received conception of a proposition, both in introducing the notion of a repetition and in allowing that we might have both p and $\sim p$ true. What I call repetition is excluded by current conceptions. Take "It was the case that Theaetetus is sitting, and then that Theaetetus is not sitting, and then that Theaetetus is sitting." The proposition that ostensibly occurs twice does not really do so, it would be said, for it is incompletely specified; "Theaetetus is sitting" must carry a time reference with it (presumably to be inferred from the context supplied by the rest of the narration), and to show this we had best represent it as "Theaetetus is sitting at time t". When the second occurrence of the words "Theaetetus is sitting" is supplied with the extra phrase "at time t'" the appearance of a repetition, which was always illusory, is avoided.

Similarly "Theaetetus is not sitting" in that proposition would not be generally admitted to be the negation of "Theaetetus is sitting" either at its first or at its second occurrence; I departed from current conceptions in allowing that we might have both p and $\sim p$ true. For $\sim p$ is defined as the proposition that is true if p is false and false if p is true. The way I introduce p and $\sim p$ as both true might seem to be harmless indeed; for it might be said that a proposition and its negation cannot be true *at the same time*, or again, the definition of $\sim p$ is sometimes expressed by saying "It is the proposition that is false *when* p is true and true *when* p is false" – and in putting "*p and then* $\sim p$" I should not be in conflict with such formulations. But these formulations do not really accord with the usual – the received – conception of a proposition. For that does not allow the variability of truth value of a proposition. But if you say that a proposition and its negation cannot be true at the same time, then if the words "at the same time" are not redundant, you are leaving it open for them both to be true at different times; but then you are admitting the possibility of a variability of truth value. The expression "at the same time" does not rightly belong in this formulation of the law of contradiction at all, if variability of truth value of a proposition is not admitted; it belongs only in the formulation "A thing cannot both have and not have the same property at the same time". Similarly, in the explanation "The negation of a proposition p is the proposition that is true when p is false and false when p is true", the word "when" only means "if", or else it ought to be disallowed, if a proposition cannot be variable in truth value. Thus my $pT{\sim}p$ can sound harmless only by benefit of slovenly formulations of the 'law of contradiction', if it is not admitted that propositions can change in truth value.

In the light of my consideration of quantification over times I can contribute one point toward the required apology. "p before q" was rendered as "a time at which p was before every time at which q". If, however, a tensed proposition is incomplete without a time specification, p and q in this formulation are not complete propositions, and so it is better to take a concrete example: "A time at which I was in Greece was before every time at which you were in Italy." Now this formulation is right for "I was in Greece before you were *ever* in Italy"; but "I was in Greece before you were in Italy" may be true, although "I was in Greece before you were ever in Italy" is false. Or again "He studied his appearance in the glass before he used the telephone" may well be a true piece of narrative; it does not at all suggest that he studied his appearance in the glass before he ever in his life used the telephone.

By the use of my "and then" and the clear explanation of repetition we can handle "before" and "after" without running into this difficulty; repetition can be excluded from our intentions, if we know what it is. If we use quantification over times it will be a complicated job to reach a formulation that is really equivalent to such a narration as "He wrote some letters before he

drank coffee" as opposed to "He wrote some letters before ever he drank coffee".[2]

VI *Events as the proper terms of the temporal relations* before *and* after

As Russell says, "we cannot point to a time itself, but only to some event occurring at that time. There is therefore no reason . . . to suppose that there are times as opposed to events."[3] It might be suggested that "before" and "after" primarily name relations between events, and that, when they are used of 'substances' or 'continuants' like the Parthenon or the Dome of the Rock, they are used in secondary senses which are explicable in terms of the primary ones. Let the letters of the alphabet stand for events, and their order represent the temporal order of the events. Then the Parthenon existed at the time of a, and the Dome of the Rock did not; both existed at the time of h, though St Peter's did not, and all three existed at the time of w. It was the overlapping existences of the buildings that give rise to the senses of "before" and "after" in which they are not converses; this can be described as we have just suggested, without bringing in "before" and "after" in any direct application to existence or to continuous states, like "being ill", "being at large", and so forth.

This move, it might be replied, will only give us a "fundamental" or "primary" sense of "before" and "after" in which they are converses, if events take no time. Russell says that "events of which we are conscious do not last for a mathematical instant but always for some finite time, however short" (p. 121). If he is right, then it looks as if exactly the same things will hold for "before" and "after" in connection with events as in connection with substances and more or less long-lasting states of things.

Or does the natural use of language prescribe conventions for "before" and "after", when they are used of the time relation between events, according to which they really are converses? Russell, indeed, says that an event is anything that can be simultaneous with something or other (p. 126n); at that rate, the existence of the Parthenon is a long drawn-out event that is still going on. This is of course not an ordinary sense of "event". But, more important, it is not the sense of "event" intended by anyone maintaining that "before" and "after" primarily name relations between events, which relations are genuine converses. A battle, a shower of rain, a visit, a collision, a conversation, a quarrel, a parting: all these would naturally be called events. Without stopping for a definition of "event" we can see that, so long

[2] In terms of T, as Mr P. Geach has pointed out to me, the difference between "p before q" and "p before *ever* q" might be expressed thus: "p before ever q" holds if and only if it has been the case that

$$(p.\sim qTq.\sim(qTp.\sim qTq)).$$

The translation of this last formula in terms of temporal quantifiers would clearly be complicated.

[3] *Our Knowledge of the External World* (London, 1914), p. 122.

as we stick to nouns, "before" and "after" as relations between events seem to be genuine converses. "The battle took place after that shower of rain" (or "after the rain"), and "The shower of rain occurred before the battle" are equivalent; and so for "That conversation took place after her visit", "Her visit took place before that conversation", and so on. The reason is that when what would ordinarily be called an event is named by a noun, and one event is said to have taken place before or after another, we naturally understand that one wholly preceded the other.

VII *Temporal clauses in which events are not named but reported*

When we turn from nouns to clauses containing the cognate verbs, the matter becomes less simple. "The battle took place after the rain" implies that the rain was over when the battle started. "The battle took place after it rained" is not inconsistent with "and it rained after the battle started". "My departure was after the quarrel and the quarrel was after my departure" sounds contradictory; whereas perhaps "I left after they quarrelled, and they quarrelled after I left" does not. "Her arrival was after their conversation" implies that no part of that conversation was after her arrival; but "She arrived after they talked to one another" does not imply that they did not talk to one another after she arrived,

It might be fairly remarked that I have to choose my examples rather carefully. Often a verb, or verb plus object, signifies an action which is a well-marked whole. Then "after p", when p contains such a verb in the simple past tense, means just the same as "after A", where A is the corresponding noun; for example, "after he read the report" and "after his perusal of the report"; or "after he spoke (in a debate)" and "after his speech". Only when a verb does not signify any well-marked whole can I plausibly make the contrast between "after A" and "after p"; for example, "after the rain", "after it rained". Further, even when I can do this, and so produce sentences "p after q, and q after p", these may well not be the most natural or clearest expression of what I have in mind. For "p after q, and q after p", where the verb of q is a simple past, may well mean "p after q, and q *again* after p". For example:

(1) They fought after it rained, and it rained after they fought – because it rained *again* after they fought.

(2) I left after they quarrelled, and they quarrelled after I left – because they quarrelled *again* after I left.

(3) She arrived after they talked to one another, and they talked to one another after she arrived – that is, they talked to one another *again* after she arrived.

To avoid the impression that this is what is meant, we need to put not "it rained" but "it was still raining"; "after they fought" not "they quarrelled" but "they were still quarrelling"; not "they talked" but "they went on

talking". And this does not give us "*p* after *q* and *q* after *p*" if we are quite strict about making exactly the same substitution for *q* both times. Further, we may say that "I left after they quarrelled" means "I left after they began quarrelling"; "they fought after it rained" means "they began to fight after it began to rain", and so on. So we just do not have genuine cases of "*p* after *q* and *q* after *p*".

Now as to the first point (the second I postpone to the next section), I should reply that when a verb does not signify a well-marked whole, if the imperfect applies, then the simple past applies. If it was raining at a certain time, then it rained at that time; if people were fighting, they fought; if they were talking, they talked; if someone was weeping, she wept, and so on. Now our topic is not usage, but whether it can be the case that *p* after *q* and *q* after *p*.

If now it should be said that all the same different events are involved in the report *q* in "*p* after *q* and *q* after *p*" in such cases, because it rained different drops of water, distinct blows were struck, fresh breath was expended in talking, new sobs were uttered and fresh tears spouted from the eyes of the weeper, then I reply that we are demanding of a single event that it should not contain parts that are events. It is not clear that this demand can be satisfied by anything but an instantaneous event; but if it can, then we can raise the same problems by introducing some event *P* (reported by *p*) that, for example, takes a shorter time than our "single event *Q*".

There is also this to be noted: when a verb – or, as is more usual, a verb plus object – does signify a well-marked whole which takes time to be complete, I cannot indeed argue "if *B* was *X*-ing then, he *X*-ed then": I cannot argue "if he was reading the report then, he read the report then"; "if he was painting the picture on that day, he painted the picture on that day". But on the other hand, in these cases the imperfect can occur in both *q*s in "*p* after *q* and *q* after *p*": "I arrived after he was reading the report, and he was reading the report after I arrived"; "I was ill after he was painting the picture, and he was painting the picture after I was ill". So here once more we have genuine cases of "*p* after *q* and *q* after *p*".

Thus, if events are allowed to take time, one way or the other in considering them we shall get a breakdown of the notion that "before" and "after" are converses, just as much as when we considered the existence of lasting states of substances or continuants; and, though it conflicts with ordinary usage, we can see Russell's point in so defining an event that even the existence of the Parthenon is an event. Russell did not at any rate think that he could handle these questions by operating with the serial converse relations "before" and "after", or "predecessor of", "successor of" such as one has for the series of natural numbers – taking them still to be proper converses in their temporal significations because here events are their proper primary field.

VIII *Are there instantaneous events?*

But was Russell right in saying that events – at least any that we experience – must occupy a finite time, however short?

He was surely wrong, and was here making the same kind of mistake as people make when they say that there is no such thing in our visual experience as a point or a line. As has often been said, there is: a line is a colour edge, and if we have a square divided into four differently coloured squares, we have a point in the middle. All that is true is that we cannot see lines or points without seeing coloured expanses whose termini they are.

Similarly, though we cannot think of an instantaneous event falling within our experience that is not a terminus of something that takes time, we can think of plenty of events that are such termini; and we may perhaps reasonably take such events as what we mean by "instantaneous" ones. In a report of such an event the verb satisfies two conditions, which I will first label and then explain. In the first place the verb is what I will call "noncontinuous"; in the second place it is what I will call "perfective". Examples are: "to arrive", "to win", "to stop".

These verbs are noncontinuous; that is, though there are uses for their continuous forms (the continuous present, the continuous perfect and, as it is called, the imperfect tense), they do not have the continuous sense. "He is winning" or "he has been winning" or "he was winning", for example, may be frequentative – "he is winning all the matches he plays in these days", "he was winning more and more often at that time" – or they may mean "he is (or has been or was) doing what naturally terminates in winning". In this sense "is winning", "was winning" do not report something the happening of which is, or was, the win. For someone can be winning and yet not win. But when a verb *F* has the true continuous sense, then if *A* was *F*-ing, *A F*-ed. For example, if *A* was seeing something for a certain time, then he saw it.

The grammatical name for the form "*A* was *F*-ing", or the endings "*-bam -bas -bat*" in Latin, and so on, is the *imperfect* tense. This name is apt as far as concerns two applications of this tense, the one we have just noticed – "he was winning" (which can be true even though he did not win) – and, second, applications such as "he was painting that picture". That the two applications are different comes out in the fact that (1) we say someone was winning but did not win and (2) we say someone was painting such and such a picture but did not finish painting it, he only partly painted it; whereas the man who was winning but did not win did not partly win, or simply fail to finish winning – he simply did not win. The designation "imperfect" is therefore apt so far as concerns these applications of the tense that is so called: but it is not apt as far as concerns the application "it was raining", "he was seeing the stain on the tablecloth", "he was living at such and such an address". In this application the imperfect signifies nothing imperfect but only continuousness; hence the imperfect, as well as the continuous present, is also called a "continuous" tense. When I say that a verb *F* does not have

the continuous sense, I do not mean that there is no use for its continuous forms.

Verbs for instantaneous events are also perfective. It might seem that being noncontinuous should be enough to determine instantaneousness; that it is not can be seen from the fact that, for example, "to know", as that verb is used of a science or a language, is noncontinuous by my explanation. It lacks the continuous sense: "he is knowing Latin" is an expression for which we have no use, and if we gave it a sense, that sense could not be the continuous present one. But "to know" is certainly not an instantaneous verb. Thus another condition is necessary, and that is *perfectiveness*. A verb *F* is perfective when (1) "*A F*-s" always implies "*A* has *F*-ed" and (2) "*A F*-s" does not necessarily imply "*A* was *F*-ing" or "A has been *F*-ing" in the continuous sense. For example, if *A* arrives at a certain place, he has arrived there; if as he makes a certain move in a game we say "he wins" we could equally well have said "he has won". Again, if "*A* sees something" is true, we could say in the same breath "he has seen it". But "he sees it" or "he has seen it" do *not* necessarily imply "he was seeing it" or "he has been seeing it". The point is clear for "see", which though perfective is also continuous, unlike "win" and "arrive". We can, however, make the same point for them: certainly "he wins" or "he has won" cannot imply "he was winning" in the continuous sense, since "to win" is not a verb with a continuous sense; and similarly for "to arrive". As we have seen, the continuous tenses of these verbs (if not frequentative) simply mean to be in process of doing what naturally terminates in winning, or arriving, not to be doing something to do which is to win or arrive.

The verb "to move" is, by contrast, a verb *F* and such that (1) "*A F*-s implies "*A* has *F*-ed" and (2) "*A F*-s" implies "*A* was *F*-ing" or "*A* has been *F*-ing", and the first point holds only because the second point holds. "*A* moves" implies "*A* was moving" or "*A* has been moving"; there is no first moment of moving, but only a last moment of rest, at any moment after which the moving body already has been moving. But this kind of implication between "*A F*-s" and "*A* has *F*-ed" is not what I am labelling "perfectiveness". (I shall be returning to this.)

In passing, we may note the contrast between "to see" and "to win". Both are perfective, but "to see" is also continuous and "to win" is not. Professor Ryle, erecting the notion of an "achievement verb", called "see" one of these and was perhaps led to this opinion partly by the perfectiveness of "to see": he apparently thought that this made it right to classify it with "to win". But as we have seen, there is also the great contrast between them provided by the fact that "to see" is a continuous verb. One can ask, for example, for how long the seeing of a certain sight by *A* went on, as one cannot ask how long a particular win went on. "Watching, I saw him for five minutes more, until he disappeared around the crag." (People are perhaps misled by the fact that "can see" is ordinary English idiom for "is seeing".)

It is also important to stress that though a language with special continuous forms is a help in expressing the truth on these matters, it is not simply a truth about such a language, and would have to be indicated by other means if we were using a language lacking special continuous forms.

Russell, then, was wrong in saying that no instantaneous events occur within our experience, because he had a false picture of what that would be like, like people who suppose that a point that could be seen would be an extensionless dot.

Yet my reader might object, "You have given a merely grammatical characterization of an instantaneous event. According to you it is an event reported by a true report whose main verb is perfective and noncontinuous. Now 'to die', for example, is such a verb. Yet there is no single distinct ground on which we can fix 'the moment of death': that is at best a legal fiction. What we want of an instantaneous event is that it should happen *at an instant*, and there are many perfective and noncontinuous verbs *F* such that it is pointless to ask 'at precisely what moment did *A*'s *F*-ing occur?' For example, scientifically if not legally, it is pretty pointless in the case of death; and if you consider 'marrying', which also appears to be a perfective noncontinuous verb, it can be seen that this grammatical characterization is extremely unsatisfactory."

To this I reply, first, that the apparent desideratum for an instantaneous event here is an illusory goal – it is notorious that when you go in for extreme accuracy you give two measurements between which, as we say, the measure falls or the thing happened – and, second, that for our present purposes the grammatically characterized instantaneous event is precisely what we want. For instantaneous events so characterized are the sort of events for which "before" and "after" name genuine converses.

IX *A beginning or an ending is always involved in* "p *after* q" *when neither clause reports an instantaneous event*

In consequence, it is perhaps still arguable that "before" and "after" are primarily relations between objects for which they are true converses: instantaneous events. On this view, uses of "before" and "after" in which these relations were not converses, like the one with which I began, would be reducible to the primary use by analysis of the sentences.

If, for example, John arrived somewhere after Tom, there is no possibility that Tom also arrived after John. But it is very implausible to suggest that for all true interpretations of "*p* before *q*" and "*q* after *p*", where the propositions are not reports of instantaneous events, we can find an analysis in which "before" and "after" join propositions *r* and *s* that are reports of instantaneous events. It may be a fact that, whatever *p* and *q* may be, if *p* and *q* are true, there was always some instantaneous event *R* that happened while *p* and while not *q* and some instantaneous event *S* that happened while *q*, such that *R* happened before *S* happened; but I do not know how such a fact could be established. In any case, there can be no good reason obliging us to think

that some such fact, or the statement that there is such a fact, is part of the analysis of "*p* before *q*", wherever "*p*" and "*q*" are reports of anything but instantaneous events; and the view is such as to need a compelling reason for holding it.

If we reject that path, can we maintain instead that if *p* before *q* then *p* and *q* themselves, if not reports of instantaneous events, must be analysable into statements each including at least one report of an instantaneous event, say *R* for *p* and *S* for *q*, such that *R* happened before *S*? Again, I can see no good reason for saying that "*p* before *q*" depends for its sense on the general possibility of such analysis for both propositions.

If we remember our Boy Scout example, however, we shall see reason to say that at any rate one *beginning* or *ending* is necessarily involved in its being the case that *p* before *q* or *q* after *p*. Now beginnings and endings are among instantaneous events, though I shall speak of them separately from each other because, as I hope to justify on another occasion, I hold the verbs "to begin" and "to stop" to be in a certain sense logical words and not verbs with a distinct empirical content: *p* before *q* if *p* before *q* began; *q* after *p*, if *q* after *p* began (the Parthenon was there after St Peter's); or after *p* stopped (the Church was there after the Holy Roman Empire); or if after *p*, *q* started (St Peter's was there after the Parthenon was there).[4]

Yet is even one beginning or ending required? I was in Rome after you were, because I was in Rome in 1955 and you were there in 1954 – however long before 1954 either of us was there, and however long after 1955 either of us stayed there. But this formulation uses dates. Can we avoid dates by saying "when I was in Rome, you had been there"? But unless this means "at *a* time at which I was in Rome, you had been there" what could it mean but "you were there before I arrived"? Similarly for "I was in Rome when you had been there": if that does not mean "at *a* time at which you had been in Rome, I was there", what could it mean but "you were there before I arrived, or I was there after you left"? The quantification over times introduced by the expression "at *a* time at which *p*" presupposes a method of identifying different times within a period for which *p*, and until that has been introduced there can be no justification for saying "*p* after *q*" without any reference to a beginning or ending.

But, it may be said, once quantification over times is introduced, we can have "*p* after *q*" without any reference to a beginning or ending; and as we saw in the case of "Jones was at large after Smith was working on his novel", the bare information "*p* after *q*" can be useful. It is useful, however, because it excludes "Smith was working on his novel only after Jones had stopped being at large". Thus it has any real significance only in contexts where there will have been, or at least may have been, stopping and starting, even though we may not know whether either *p* or *q* stopped or started. To say of two

[4] The expression "*p* began" is to be understood by knowing how to construct sentences that would be instances of the variable proposition "*p* began". If for *p* we substitute "it was night", then for "*p* began" we substitute "it began to be night". Similarly for "*p* stopped".

beings conceived as everlasting, like Aristotle's planets, that one existed after the other would be quite empty.

X *Summary*

If *q* does not report an instantaneous event, "*p* before *q*" implies that *q* began, and is equivalent to "*p* before *q* began". "*p* after *q*" can be reasonably used even when neither clause reports an instantaneous event: but it is pointless if we know that there is no question that "*p* began" or "*p* stopped" or "*q* began" or "*q* stopped" should any of them be true.

Let us assume that "after *r*" is unproblematic if *r* reports an instantaneous event or the beginning or ending of something.

Then:

(1) Suppose *q* reports an instantaneous event. Then "*p* after *q*" is unproblematic, but is consistent with "*p* before *q*", unless *p* also reports an instantaneous event.

(2) Suppose *q* does not report an instantaneous event, but *p* does. Then "*p* after *q*" means the same as the unproblematic "*q* before *p*", but is compatible with "*q* after *p*."

(3) Suppose neither *p* nor *q* reports an instantaneous event.

Then *p* after *q* if:

(a) *p* after *q* began, without prejudice to whether also *p* before *q*; or

(b) *q* before *p*; or

(c) *p* after *q* stopped, without prejudice to whether *q* began, or to whether *p* also before *q* stopped.

Now we can understand – but also we can see reason to resist – the inclination to suppose that on analysis "before" and "after" will turn out to be relations between events. The stopping or starting of something is an event and does come in, we may say, as a term of "before" and "after" relations wherever we have "*p* before *q*" or "*p* after *q*" even though neither *p* nor *q* itself reports an instantaneous event. But that is all. We shall go wrong if, moved by the feeling that events are the proper terms of these relations, we insist on looking always for two events wherever we have *p* before *q* or *p* after *q*.

18 Subjunctive Conditionals

Prefatory Note

The following piece was written as a try-on and in reaction against standard arguments that the "if" of the subjunctive conditional is non-truth-functional. It seems to me still to be of some interest, partly for its rejection of those arguments which are indeed surely wrong, and partly because one can go surprisingly far with the assimilation. But the piece may be mistaken. It has been argued to me, for example by Fabrizio Mondadori, that the equivalence I asserted between:

> If there were a meat shortage, then if meat prices were low there'd be governmental control

and

> If there were a meat shortage, then either prices would not be low or there'd be governmental control

can't be correct, because the second but not the first could be said to be *verified* by a meat shortage together with high prices and government controls. There is something in this; but it is uncertain.

Further, G. H. von Wright remarked to me that the analogue of the entailment of $(p \& q) \supset r$ by $p \supset r$ does not hold for subjunctives. Certainly it does not hold generally.

My principal error consisted in thinking that, if an "*if* . . . *then* . . ." is neutral in respect of the particular sort of connection that justifies its assertion, then that neutral "*if* . . . *then* . . ." must be truth-functional. I assumed that there must be a constant completely specific logical form common to all subjunctive conditionals. I now think this is no more the case than it is for indicative conditionals. For these, the specifications of logicians have been directed towards replacing an indeterminate form by determinate ones. The indicative conditional is characterized by *modus ponens* and *modus tollens*, but once we have said that we have left much undetermined, e.g. about the truth-conditions of conditionals. People sometimes speak of the material conditional as a minimum conditional, but it is so only from a certain point of view. We might think of a minimal conditional as one that is *only* characterized by *modus ponens* and *modus tollens*, and this might be thought of as generic or skeletal. Context may then prove that here the "*if* . . . *then* . . ." has a more specific form.

Might not the same hold for subjunctive conditionals? One might enquire

An article in *Ruch Filozoficzny* (Warsaw), 33, 3/4, (1975), with a new prefatory note.

what the corresponding indicative conditional would be: a truth function or an entailment or a statement of a connection of nature? If the indicative conditional is one of entailment, then presumably the subjunctive is too: "If *m*'s digits were to add up to nine or a multiple of nine, *m* would be divisible by nine." And similarly, where the connection is a connection of nature.

Now suppose the indicative is truth-functional. Can we give an example? It might be promising to take "If Tom is out, John is in", which I might state because I had given an instruction, on that occasion, that Tom was on no account to go out without John's being sure to be in. If this is truth-functional, then won't the subjunctive "If Tom were out, John would be in" be equally so?

But it may be said: Do you mean the indicative to imply "If Tom is out, and I gave no instruction on the matter, John is in"? Or even, "If Tom is out and John is out too, then John is in"? It might be argued that the subjunctive is not truth-functional because "If Tom were out and I had given no instruction in the matter, John would be in" surely doesn't follow from "If Tom were out, John would be in". Even if someone suggested a "no matter what" intention about the statement, we would hardly have "If Tom were out and John were out too, John would be in". Whereas in the indicative we really can say that "If Tom is out and John is out too, then John is in" follows from "If Tom is out, then John is in", if this is truth-functional. For this is nothing but an instance of $(p \supset r) \vdash (p \& q) \supset r$, which is completely harmless. That it is harmless, however, proves that we haven't got an example with a real application. And it might be maintained that, given a really practical example of truth-functional if–then, it could be cast into the subjunctive without detriment to its implications.

There would be an objection to this if subjunctives had a point which ruled it out. But I do not think they have.

ORIGINAL ARTICLE

In discussing subjunctive conditionals it is common to speak of the truth or falsity of the antecedent and consequent. At first sight it seems rather pedantic to correct this way of speaking and to insist on saying: "the truth or falsity of their translations into the indicative".

Subjunctive clauses, however, are not propositions, though they exhibit the predicative connection of expressions; e.g. the predicate "struck" has the subject "this match" in "this match were struck". The clause is called false if the predicate does not hold of the subject. But to say that is merely a circumlocutory way of formulating the indicative proposition that corresponds to the subjunctive clause. It therefore remains correct – though perhaps pedantic – to say that we cannot strictly speak of the truth-value of the subjunctive clause; what is meant is perforce the truth-value of its translation into the indicative.

The currently received view assumes that what is in question here is the

relation between two indicative propositions. When it is said that the "if" cannot be truth-functional, it is assumed that the translation of the whole sentence:

If this match had been struck, there would have been flame

into:

If this match was struck, there was flame

is correct.

Some would deny this on the ground that the subjunctive is 'counter-factual' but, as has often been remarked, subjunctive conditionals are not as such counter-factual. They are often used for example in seeking historical explanations of what did happen. Mammoths (it is said) are found whole with fresh buttercups in their mouths in the Siberian ice. How did this come about? If such and such had happened then . . . and then mammoths would have been found whole, etc. Thus we need not add negations of the antecedent and consequent when we represent what is assumed to be a correct translation in the well-known arguments that this "if" is not truth-functional.

The argument may then be rendered: "The translation into the indicative is correct, but certainly for the subjunctive conditional it matters for the truth of the whole what consequent is connected with the antecedent. If the connection of the indicative propositions in the translation is truth-functional, then it does not matter for the truth of the whole what consequent is connected with the antecedent if the antecedent is false. Therefore the connection is not truth-functional."

But it might be replied: "The proposition offered as a translation is one for whose truth-value it is indifferent what consequent you have, if the antecedent is false. And certainly for the subjunctive conditional it matters for the truth of the whole what consequent is connected with the antecedent. Therefore the translation is incorrect."

To this is might be objected: You are assuming that the "if" in the translation is truth-functional. But suppose it is not?

Of course we then need an account of a non-truth-functional "if" here, and reason to suppose it occurs. Note that the "if" of subjunctive hypotheticals can't be offered in evidence of this; the argument that it is not truth-functional relies on speaking of the truth-values of the clauses and this assumes that the translation is correct, which is the point at issue.

One thing that could be said: as "if" is commonly used, in *many* contexts it is natural to say that a conditional is neither true nor false when the antecedent is false. Someone asks: "Will she kiss him if he wins?" If he does not win, it may well seem sensible to brush aside the query: "But what was the true answer to that question?"

This however, even if acceptable, would be no great contribution. It would indeed give us an "if" that is not truth-functional, but this hardly

affects our criticism of those who argue that the "if" of subjunctive hypotheticals is not truth-functional. We may state the criticism thus:

> Leave the clauses in the subjunctive, and you cannot speak of the falsehood of the antecedent, but it does matter for the truth of the whole what consequent is connected with a given antecedent. Translate them into the indicative and, though you can now speak of the falsehood of the antecedent, it is indifferent what the consequent is if the antecedent is false. For then the whole is either indeterminate or true, whatever the consequent. Which it is, will depend on whether the "if" is truth-functional (as it often is) or not truth-functional in the way explained (as *perhaps* it sometimes is).

"But surely not!" it may be exclaimed. For though "Will she kiss him if he wins?" may sometimes reasonably be said not to have a true answer if he doesn't win, "Will she stand on her head if he wins?" may well, in just the same circumstances, reasonably be given a negative answer. The negative answer may be *certain*. Or again, two people dispute, saying:

If she comes, he will be angry
If she comes, he won't be angry

and even if she doesn't come, one may be right, may know better than the other. One of them may know which that man *would be*, angry or not, if she *were to* come.

That is reasonable indeed – but it only means that, if the indicative has a *correctness* which is not a function of the truth-values of the component propositions, that correctness is *grounded* upon the *independent* truth of the subjunctive hypothetical in question. And that is enough to show that the indicative is not a translation of the subjunctive. With that, this whole argument that the "if" of the subjunctive hypothetical is not truth-functional falls to the ground. For the argument talked of the truth-values of the clauses, and that assumes that they are translated into the indicative.

It has hitherto been taken for granted that the clauses connected in subjunctive hypotheticals are propositions, and consequently it has been argued that the connection is not truth-functional. I wish to say that the clauses are not propositions, but I shall argue that the connective *is* the truth-functional "if": that is to say, that the "if" is not to be regarded as different from the "if" that we most generally get connecting propositions and predicates. I say "most generally" because we say "if . . . then . . ." in a context like "If the potato crop was halved by blight then either there was a shortage or there had been excess production before," without any feeling of zeugma or gear-change, though of course the relation of shortage to potato blight is quite different from that of previous excess production. Someone might produce the conditional sentence with a view to showing that the antecedent was false, it being provable that neither disjunct in the consequent is true. So the "if . . . then . . ." is as it were *neutral* in respect of the particular connection which may justify it. Such a neutral connective is truth-functional. The only non-truth-functional "if" that there is so far any reason to introduce is non-

truth-functional only to this extent, that, e.g., "If he wins she will kiss him" may perhaps sometimes reasonably be called indeterminate in truth-value if he does not win. But there is no place for such an "if" in subjunctive conditionals. Not because the whole must have a truth-value; on the contrary, "If he wins she will kiss him", would be reasonably called indeterminate when he does not win precisely in cases where "if he were to win she would kiss him" could be called indeterminate. But, just because we can't speak of the falsehood of the antecedent unless it is translated, there is no room for calling the "if" in that subjunctive hypothetical non-truth functional on the ground that the truth-value of the whole is indeterminate *when the antecedent is false*.

When I say that in "If Jones were *F*, Tom would be angry" neither "Jones were *F*" nor "Tom would be angry" is a proposition. I am pointing to the fact that they cannot be taken out of their contexts and significantly asserted. But this is not what it may seem, an 'ordinary language' argument. For consider "Smith wondered whether Jones were *F*". I see no objection to calling "Jones were *F*" a proposition as it occurs in that context. We can remove the oddity of calling *this* "Jones were *F*" a proposition, or of supposing it removed from that context and asserted, by translating; the proposition in question is of course "Jones is *F*", for Smith wondered: is Jones *F*?

The point is that, as the foregoing argument has been designed to show, the corresponding translation of the subjunctive hypothetical does not yield an equivalent proposition. For either (1) it yields a proposition which must be true or indeterminate If Jones is not *F*, whereas "if Jones were *F*, Tom would be angry" is *not* proved true or indeterminate by Jones' not being *F* or else (2) if the indicative hypothetical is not made to be either true or indeterminate just because Jones in not *F*, that must be because of the independent falsity of the subjunctive hypothetical. In either case, then, the whole indicative proposition is not equivalent to the subjunctive one.

Nevertheless the question does remain: *is* the "if" of subjunctive conditionals truth-functional? Certainly not, if truth-functions are explained as functions whose arguments and values are truth-values.

However the truth-connectives connect not only propositions, but also predicates. $(Ex)Fx \supset Gx$ and $(x)Fx \supset Gx$ cannot quite generally be explained as disjunctions and conjunctions of singular propositions. Here we have on the one hand a method of construction, and on the other, rules connecting generalized propositions with (possibly suppositious) singular propositions. May we not similarly have a method of construction of subjunctive propositions, and rules connecting subjunctive propositions with the indicative propositions corresponding to the subjunctive clauses? We start with singular propositions *Fa*, *Gb*, and from these we form open sentences *Fx*, *Gx*. From these we form generalized propositions $(Ex)Fx$; $(x)Gx$ by binding the variable. Equally I suggest we may form open clauses $F'x$, $G'x$; the prime is an operator on predicates: We read: "It would be that *Fx*" or "*G* would hold of *x*" or "*x* would be *F*". I speak of open clauses, not open sentences, because these are open incomplete sentences: substituting a name for the variable, or binding it with a quantifier, produces an incomplete sentence, i.e. a clause which is not a sentence.

We maintain the truth-functional equivalences – e.g. between $Fa \supset Fb$ and $\sim(Fa \sim Fb)$ – for the corresponding open sentences. When we construct generalized sentences, these equivalences still stand: $(Ex)(y)Fx \supset Gy$ is equivalent to $(Ex)(y) \sim Fx \vee Gy$. That is indeed the rationale of 'maintaining' them for the open sentences. It is also the *ground* for calling the connectives the same as those that connect singular propositions truth-functionally.

Now if these equations work for the "if. . . then . . ." that we get in subjunctive conditionals, we have all the ground that there could be – and the same ground as there is with quantified propositions – for identifying it with the truth-connective. So we want to investigate whether the equations do work.

Before starting the enquiry, it will be of use to notice some sentences which occur as single subjunctive predications but which are not of interest to us. Some are deprecatory, as "John would be in now", which may be a deprecatory way of asserting that John is in now. Or it might be called elliptical for "I think John is in now". Another quite different case is the use of a subjunctive to describe how things typically were in some past situation: "There would be a lot of noise". "And then there would be no fire". I don't know the rationale of this, but it is evidently *equivalent* to an indicative taken frequentatively.

We might want to say that "That wouldn't happen" is elliptical for "Whatever else were to happen, that wouldn't", because the expansion doesn't modify what is said; but one might equally say that the expansion was merely redundant. Now whatever one wants to say about proper free-standing single-clause subjunctive sentences, it is clear that they are *relatively* rare and somewhat problematic in sense. That is the real reason for saying they are elliptical. Free-standing disjunctions, and still more so conjunctions, in the subjunctive mood are also rare; they may seem unnatural, or sometimes archaic. I will come to them. But I will first test my hypothesis on examples where the subjunctive conditional is itself the antecedent or consequent of a larger subjunctive conditional. Here, these difficulties don't arise.

Our topic is heavily entangled with idiom: at least, it is so in English. In English we have the true subjunctive, e.g. "If . . . were . . ." and the periphrastic subjunctive " . . . would be". In more archaic language there was less use of the periphrastic subjunctive. In examples I shall use whichever is idiomatically natural in current English, but I wish to note that logically speaking the change from one to the other is of little significance, and is of none for our purposes.

We begin with the equivalence of "If *p* then *q*" and "Either not *p* or *q*".

(1) If, if John were not in Tom would be in, I could help you

is equivalent to:

(2) If either John were in or Tom were in, I could help you.

And:

(3) If there were a meat shortage, then if meat prices were low there would be government control
(4) If there were a meat shortage, then either meat prices would not be low or there would be government control.

Note that for our present enquiry we need not give consideration to natural assumptions as to signs, causes and effects in our examples. For example the last clause in the immediately preceding example might be: "there would be a black market". "If your husband had to pay £50,000 damages, what would he be?" may expect an answer like "Furious", but get one like "A rich man". The tense of the subjunctive clause has some bearing on this matter, e.g. one might put "There would *have been* government control in (3) and (4) (making some natural assumptions). Some complicated periphrases may have point. E.g. consider "If Mary were not being tactful, John would be angry" and try to construct the contrapositive. "If John were not angry, Mary would be being tactful" seems not quite to succeed: rather we are constrained to put, e.g., "Were it the case that John was not angry, then it would be the case that Mary was being tactful". The reason is that the former, because of its content, perhaps imposes the suggestion that John's anger would affect Mary so as to prevent her from being tactful. The original proposition is ambiguous in its suggestion: perhaps John would be saved from being angry by Mary's tact, or perhaps Mary's failure to be tactful would *make* John angry. I suppose that all these things are examples of what Grice calls "implicatures". Our final effort at constructing a contrapositive succeeds by being laboriously neutral and unsuggestive. Now I wish to put all this sort of thing on one side: if an equivalence fails to be convincing *only* because of suggestions as to cause and effect, temporal order, sign and thing signified, that does not fault the equivalence for the purpose of the present enquiry.

So far, then, *if F'a then G'b* seems to be equivalent to *not F'a or G'b*. That is to say, they seem to be equivalent when the two subjunctive clauses so connected occur as either the antecedent or the consequent of a conditional. But is there still such an equivalence when *if F'a then G'b* is the whole sentence?

Here we have to observe, what we have already observed in practice, that the periphrastic subjunctive with "would" has ousted the true subjunctive "were" in conditionals, except in the antecedent, but that this is a mere idiom. We do not put:

I were not angry, or else you would be cowardly

but:

I would not be angry, or else you'd be cowardly

when testing for the equivalence of a sentence composed with "not . . . or" to:

If I were angry you would be cowardly.

The equivalence works, as also in the protest: Such sentences would not occur as equivalent to subjunctive conditionals, or else it would be the merest trick of idiom.

There is a mixed form:

No stranger approached, or the dog would have barked.[1]

This may be seen as derived from a subjunctive disjunction which goes best in an archaic form: "No stranger had approached, or else the dog had barked" with one arm of the disjunction replaced by the indicative. For the mixed pronouncement would be based on:

If any stranger had approached, the dog would have barked.

There is of course some reason for the inclination to call subjunctive conditionals "counter-factual", as if this were a form: the knowledge that *not Fa* or that *not Fb* is often the background for saying *if F'a then F'b*. The mistake is only to think that counter-factualness belongs to the account of the subjunctive conditional as such. Counter-factualness is obviously a semantic, not a syntactic, property.

Let us now test for the equivalence of *If F'a then G'b* and: *not both F'a and not G'b*.

I should be able to help, unless both John were in and James not in

is equivalent to:

If if John were in James were in, I should be able to help.

Trying the consequent for the same result:

If John were poor, then not both Tim would be in France and James not on the farm

comes up as rather unidiomatic. The idiomatic form which comes to the same thing, would be:

If John were poor Tim wouldn't be in France *and* James not on the farm

and these are equivalent to:

If John were poor, then if Tim were in France, then James would be on the farm.

[1] An example suggested by a slightly different example given by Bernard Mayo, 'Conditional Statements', *Philosophical Review* (July 1957).

So much for these apparent equivalences in the antecedent or the consequent. When, as with "or" we look to see whether freestanding *if F'a then G'b* has an equivalent: *not both F'a and not G'b*, we note the same idiom as occasioned our translation of the stilted proposition about John, Tim and James:

John wouldn't be in *and* James not

which means:

Not (John would be in and James not).

Incidentally, if we reflect on the saying:

It never rains but it pours

we can discern an archaic and rustic "but", used only after negations, and meaning "and not". The word "without" has a similar use, though not in what is called good English. The predilection for an antecedent negative – partial for "without" so used, and absolute for "but" so used – is an objection to rendering these words "unless" which has no such requirement; all is explained, when we think of the equivalence of *if p then q* to *not (p and not q)*. Thus our sentence means:

It never rains and does not pour

i.e.

Always, if it rains, it pours.

Similarly someone might say:

He wouldn't come to this town but he'd come and see me

i.e.

Not: He would come to this town and not visit me.

Thus the equivalences of material implication also hold for the "if. . . then" of subjunctive conditionals. Similarly exportation holds:

If John were in and Tom were out, Father'd be left alone

is equivalent to:

If John were in then if Tom were out, Father'd be left alone.

And also contraposition. This fact may be hidden from people by the idiom by which the periphrastic "would" is usual in consequents. Using an older idiom

Were *T* alive, then were *W* dead

contraposes to

Were *W* not dead, then were *T* not alive.

And we may use the same translation rule as before for more modern constructions.

Thus *if F'a then G'b* displays all the equivalence characteristic of $p \supset q$. Now there is one aspect of the equivalence between "if . . . then . . ." and "not . . . or" . . . which we have not considered. The equivalence means that *if not p, then q* is the equivalent to *p or q* and it works for subjunctives when the whole phrase is itself the antecedent – or the consequent – of a conditional. But what if *if not F'a then G'b* stands alone as an assertion? Is it then equivalent to *F'a or G'b*? The difference between this and the examples previously considered is that here the disjunction contains no negative.

Several things impede the consideration of this. We have seen that what in superficial grammar is a single subjunctive predication may stand by itself:

John would be in now.

This is a form of statement that John is in now. Then

Either John or Tom would be in now

might be taken as the disjunction of two deprecatory statements. If that did not matter we'd have our example. But it may seem that it does matter, for if this is a disjunction of propositions, each has a truth-value just as it stands.

Leaving this example, we may have:

This would float – or ships would sink.

where the equivalence holds. But it may be said that the antecedent is elliptical for:

This would float if it were put in water

and similarly in:

This water would boil at such-and-such a temperature, or else the pressure would be above such-and-such

it may be said that "at such-and-such a temperature" has the effect of an if clause "if it were raised to such-and-such a temperature". We are looking for, so to speak, pure examples. We can find them; they are curiously oracular:

Nylon stockings would fail – or nylon spinners would
Napoleon would have been a conqueror – or else have died young
God were faithful – or our cause were lost

These are clearly enough equivalent to the more ordinary forms: *if not F'a, then G'b*.

We noted some occasions of hesitation about "Either Tom or John would be in". There are not such occasions with:

Tom wouldn't be out *and* John away:

which is the idiomatic form for:

Not both Tom would be out and John away.

But now, what about the equivalence of $p.q$, $\sim(\sim p \dot{\vee} \sim q)$ and $\sim(p \supset \sim q)$? Is a corresponding equivalence going to hold between *F'a and G'b, not either not F'a or not G'b* and *not: if F'a then not G'b*? And (a connected question) what about a free-standing conjunction of simple subjunctive clauses?

We find that we have the equivalence when these structures occur as antecedents; this may be seen in the following examples (for naturalness "unless" is put instead of "if not"):

(1) If John were in and Tom were out, Father'd be alone.
(2) Unless if John were in, Tom were not out, Father'd be alone.
(3) Unless either John were not in or Tom were not out, Father'd be alone.

(1) If he were sick and tiresome, she'd leave him.
(2) Unless if he were sick he were not tiresome, she'd leave him.
(3) Unless he either were not sick or not tiresome, she'd leave him.

(1) If it were red and granular, it would be poison.
(2) Unless if it were red it were not granular, it would be poison.
(3) Unless it were either not red or not granular, it would be poison.

In each of these examples (each of which suggests a different type of 'connection') (1), (2) and (3) are all equivalent. Now suppose that *F'a and G'b* is a consequent. From:

If Father had made a will, Jim and Michael would have been disinherited

we get as equivalent:

If Father had made a will, neither (not either) Jim nor (or) Michael would *not* have been disinherited.

What about the equivalence of ". . . and . . ." to "not (if -- then not . . .)"? This is at first sight rather a stumper, partly because of the unnaturalness of the proposed sentence:

If Father had made a will, then not: if Jim had been disinherited, Michael would not have been.

There might seem to be a natural way of saying this, e.g.:

If Father had made a will, then we wouldn't have had Michael not being disinherited, if Jim was disinherited.

But this seems to leave it open that possibly neither would have been disinherited and only to exclude the disinheritance of one *without* the other. What we want is rather:

. . . We wouldn't have had this: if Jim had been disinherited Michael would not have been.

and we need to interpret the sentence after the colon as we have already argued we should interpret the subjunctive conditional occurring as consequent, getting:

. . . either Jim would not have been disinherited or Michael would not have been.

In conclusion, consider the following question and answer:

Would this hall hold 2,000 people?
It would hold 2,000 *and* stifle them.

19 "Under a Description"

When I introduced the phrase "under the description" as a tool in the philosophy of action,[1] I thought it something that couldn't be called in question or misunderstood. Subsequent history has educated me. Here is a list of the challenges or misunderstandings that I have noted, with some comments on each:

(1) It is supposed by some that "*x* under the description *d*" is the form of a subject phrase. This of course raises the question what sort of object or entity (distinct from an A?) an A-under-the-description-d may be. But "Under the description 'putting the book down on the table' my action was intentional, though it was unintentional under the description 'putting the book down on a puddle of ink'", has as subject simply "my action" and as predicates "[un]intentional under the description '. . . '". As some people have observed "under the description" is '*qua*', or Aristotle's 'ᾗ' in modern dress. Aristotle too observes that the phrase 'ᾗ . . .' belongs to the predicate, not to the subject (*Prior Analytics*, I, Chapter 38). There aren't such objects as an A *qua* B, though an A may, *qua* B, receive such-and-such a salary and, *qua* C, such-and-such a salary.

(2) In view of (1) the recognition that some single object may be φ *qua* B, (or φ under the description "B") and not φ *qua* C has nothing to do with Leibniz' Law. I once wrote that, since most human actions consist in physical movements, it couldn't be that the actions were free (not necessitated) though the movements were determined (necessitated).[2] Dr. A. Kenny invented a reason why I thought this, namely adherence to Leibniz' Law; and (in the first series of Gifford Lectures in which he was one of the lecturers) he criticized me for it, as one can't safely appeal to Leibniz' Law in modal contexts. When I complained of this, he apologized in his next book but one,[3] saying that his criticism was indeed unjustified, as I was known to have rejected Leibniz' Law in my published writings! The truth is that I have never published a word about Leibniz' Law. The 'rejection' consisted only in my saying such things as that an action may be intentional under one description and not intentional under another. But this is no more a rejection of Leibniz' Law than it is to say that Socrates is taller than Theaetetus and not taller than Plato.

(3) I have on occasion stared dumbly when asked: "If one action can have many descriptions, what is *the* action, which has all these descriptions?" The question seemed to be supposed to mean something, but I could not get

[1] *Intention* (Oxford, 1957), p. 11.
[2] *Causality and Determination*, chapter 13 of this volume.
[3] *Will, Freedom and Power* (Oxford, 1976). p. 153.

From *Noûs*, 13 (1979)

hold of it. It ought to have struck me at once that here we were in "bare particular" country: *what* is the subject, which has all these predicates? The proper answer to "What is the action, which has all these descriptions?" is to give one of the descriptions. Any one, it does not matter which; or perhaps it would be best to offer a choice, saying "Take whichever you prefer".

(4) On the other hand, some writers have apparently embraced some rather strange idea of 'under a description'; have at any rate used the phrase in a most mystifying way. Thus J. Bogen's interesting article "Physical Determinism" is marred by the use of the expression "done under the description . . ." and "performed under the description . . .". This indeed opens one's eyes to what some people may have found objectionable about the expression. What can be meant, for example, by saying that something happened under one description, but not under another? No natural sense suggests itself for "happening" or "being done" or "being performed" together with the phrase "under the description *d*". At best, the phrase seems redundant – one might say: What happens happens under every description that is true of it! Whereas – and this was the point of the expression – there may be descriptions that are true of a happening, though the happening is not known, or willed, or desired, or explained under those descriptions. It is usually to the point to say that something was not intentional under the description *d*, only when the description *d* is true of it. Certainly one would never say "It was unintentional under that description" unless the description was true of it. However, phrases such as "under a description", "under the description *d*" sometimes occur in hypothetical contexts or in a generalized way, as in: "A reason for an action is a reason for it under some description". Here the description must be supposed – by whoever has the reason – to be one that will apply to the act if it is done.

(5) This readily leads me to the next difficulty. Animals that have no language can have intentions too: how then, it is asked, can it be right to say that an intention is always 'under a description'?[4] Again I found the objection puzzling: another non-reason. But I suppose that one who offers it must be taking "it was intentional under this description, not that" to imply that the first description is in some sense written into something inside the agent. And the form of expression I have used at the end of (4) might perhaps suggest something of the sort: it sounds as if the agent had a thought *about* a description. But now let's suppose that a bird is landing on a twig so as to peck at bird-seed, but also that the twig is smeared with bird-lime. The bird wanted to land on the twig all right, but it did not want to land on a twig smeared with bird-lime. If it landed on the twig in order to peck at the bird-seed, can't we say it took landing on the twig to be a way of getting into a position to peck at the bird seed? We can, if we can say that a bird thinks it can escape into the open by flying towards the daylight that comes through a glass barrier. This way of talking does not presuppose that the bird has any thoughts about descriptions. If there is a difficulty, it concerns ascribing

[4] Cf. D. Bennett, "Action, Reason and Purpose", *Journal of Philosophy*, 42, 4 (1965).

those other thoughts to the bird; it is about passing from the bird's intentions or aims, to the ascription of belief to it. But someone who says the bird's action was intentional (or voluntary) under one description, not under the other, need not enter into *that* dispute at all. "It took landing on the twig to be a way of . . . but not a way of . . ." is merely a rather roundabout way of saying that, e.g., the bird meant (wanted) to land on the twig, but not to land on the bird-lime. Landing on the twig was landing on bird-lime – we aren't considering two different landings. So, if we form definite descriptions, "the action (then) of landing on the twig", "the action (then) of landing on a twig with bird-lime on it", we must say they are definite descriptions satisfied by the same occurrence, which was something that the bird did, but under the one description it was intentional, under the other unintentional. That the bird is not a language-user has no bearing on this.

(6) If one says that one and the same action (or other event) may have many descriptions, it is sometimes supposed that this must be said in the light of a theory of event-identity. Now this appears to me no more true than that one can only say one and the same man may satisfy many different definite descriptions in the light of a theory of human identity. There may be different theories of human identity, yielding different results in curious describable cases. But what would we say of a theory which grants that a certain man, Dickens, wrote *David Copperfield* and *Bleak House* and that only this Dickens wrote *David Copperfield*, and only this Dickens wrote *Bleak House* – but does *not* grant that "The author of *David Copperfield*" describes the same man as "The author of *Bleak House*"? (As Hughes and Cresswell say that *the top card of the pack* is a different card from any of the fifty-two.[5]) We'd say that it is a non-starter: any *theory* of human identity has got to fit in with the correctness of calling the author of *David Copperfield* the same man as the author of *Bleak House* (subject to astonishing literary discoveries, which are not our concern here). To say that one must have a theory of human identity if one says that, or that one needs it in order to justify saying that, seems to be absurd. A theory, I suppose, will at least (a) determine answers in obscure or borderline cases; (b) give an interpretation of known facts. It may sometimes also correct generally received statements in the light of new knowledge of a general scientific nature. The Dickens case is affectable only under (b), i.e. not in such a way as to result in a change in the truth-value we assign to the identity statement.

For this reason I always balked at the question "What is your theory of event-identity?" or "What theory of event-identity lies behind saying that (in the imagined case) putting the book down on the table and putting it down on an ink puddle were the same action?" Any 'theory of event-identity' had better yield this result: it itself is not a theory or part of one.

But now one may find (or invent) new verbs which are equivalent to old ones with certain complements, or which, with certain complements, are equivalent to old ones. A difference of verb, then, (such as, say, that between

[5] Hughes and Cresswell, *Modal Logic* (London, 1968), p. 147.

'pierce' and 'stab') doesn't have to make a report into a report of a different event, if the omission of part of a complement does not have to.

As Davidson has put it, all that he (or I) meant by speaking of many different descriptions of one action is, e.g., that the executioner of Charles I, having taken his head off, did not have to add any further performances, namely of killing and of executing, to make his act one of killing and executing. What we meant, in short, is something that isn't a philosophical thesis at all, and which no one denies. What is under dispute is whether to *speak of* many different actions – perhaps as many as there are (possible?) different descriptions, perhaps fewer than that – in the circumstances where I (and Davidson) speak of only one. For us the question "How many landings did the bird make?" has a straightforward sense; now suppose our answer is "Just two" – what we express in that way will (by others) have to be characterized differently, if landing on the twig and landing on the limed twig are *eo ipso* different landings. Or – in the case where I would call them the same landing – will others call them the same landing but not the same event or action? How many battles were fought at Waterloo in 1815? There seems to be not much doubt who is in the terminological difficulty here.

(7) Alvin Goldman is a critic who does not accept the sort of result that I say any 'theory of event identity' had better yield. He thinks that if I said "Hello" loudly on a given occasion, my saying "Hello" and my saying it loudly were two different actions. This puts the conception that he wishes to explain under the title "identity and difference of acts" right outside anything that I have been considering; the topic is so altered that, if it were not for certain criticisms that he makes, I would have supposed that we passed one another by, and would not have recognized what you could exactly call a *disagreement* between us. What he appears to mean by an "action" is a supposed correlate (other than what I call an action) of an individual action-concept: merely saying "Hello", for example, on a particular occasion, *in abstraction from* any degree of loudness or tone of voice in which one actually says "Hello". I would indeed expect such events as these to be as little viable as the intensional objects discussed by Quine in 'Reference and Modality'[6]: but, even supposing that one can maintain a theory of these 'actions', they aren't what I am talking about.

However, Goldman has some criticism of what he calls the "identity thesis". It is noticeable that he begins his book[7] by saying "Suppose John (1) moves his hand, (2) frightens away a fly, (3) moves his queen . . . [etc. up to (6)]. Has John here performed six acts?" He then ascribes to Davidson and me the answer that John performed only one act in this case. But naturally I can have no views on the case as described. It reminds me of the would-be sceptical question: 'can we know that other people see?' If one asks which other people, it turns out to be the sighted ones that are meant. Do I say John

[6] In the second edition of *From a logical point of view*.

[7] *A Theory of Human Action* (Englewood Cliffs, New Jersey, 1970)

performed one act? Well, in which case? We quickly guess from his discussion that Goldman means: in a case where he *did* do all that just by what it is rather natural to call one act. He wasn't for example playing half a dozen games of chess at once, in one of which he moved his queen while in another he checkmated his opponent, simultaneously frightening a fly away by blowing a raspberry. Etc. But the case as actually specified leaves it open whether John did all those things by 'one' act, specified as "moving his hand", or by any number up to six (or more). When it is taken as Goldman intends, the identity of the act is assumed in giving a lot of different descriptions.

Most of Goldman's objections are rather frivolous – a man paying a collector of two-dollar bills a debt of two dollars with a two-dollar bill, when he could have given him two dollar bills. Naturally I have no difficulty here. Under one description what he did was something he was under no obligation to do and it was nice of him . . . (see (1)).

Again, Goldman dodges about between ostensible singular terms – terms which, whether proper names or definite descriptions, are apparently supposed to function as designations of individual acts – and expressions like "He checkmated his opponent by moving his queen". Since this is different from "He moved his queen by moving his queen", Goldman thinks he has pointed to an asymmetric and irreflexive relation between acts, which shows their non-identity. But one might as well argue that because the US President is the US Commander-in-Chief by being President, while the Commander-in-Chief is not the President by being the Commander-in-Chief, the President and the Commander-in-Chief can't be the same man. (The relevant comment *here* is that definite descriptions in predicative positions are predicates.) The question whether two phrases are true of one and the same action can't be settled by pointing to their non-substitutivity if they are anyway not singular designations. If I say, e.g., "In this position moving the queen *is* delivering mate", I am not uttering an identity statement but characterizing a type of move as a mating move. If N did move so and someone says that his then act of moving the queen was his act of mating, this is a rather stilted way of talking; but the obvious thing for him to mean is that it was the mating move of the game. Once again, this is predicative; we should not be led by the definite article to speak (yet) of an identity proposition any more than in the case of the President and the Commander-in-Chief. "Mate-delivering move" is *ex hypothesi* a true description of the move that the player made, which therefore satisfies the definite description: "The mating move of that game." But he is certainly not being supposed to have made two moves, and that move also satisfies the definite description "the move [or, to be supercareful, the last move] of queen to . . . in that game". If two moves are not in question, then we have an identity: "the mating move" and "the move of queen to . . ." describe one and the same move. So someone who says the player performed two *acts*, though only one move, must proceed carefully, neither confounding the acts nor dividing the move. It is not so far clear what the point is, of distinguishing in this fashion between the acts and the move when each of the acts being considered is a move.

Some people don't like recognizing events or actions as individuals. There is, they point out, no answer to the question "How many actions did you do this morning?" or "How many events have taken place in this room in the last hour?" This, however, merely shows that neither "action" nor "event" is much use as a count-noun, but there are many count-nouns that apply to actions and events, e.g. "death", "kiss", "explosion". Just the name holds for *material objects*, or *things*. "How many things are there in the room?" is unanswerable unless contextually specified. A difficulty about how events or actions may be uncountable, while deaths which are events, and kisses which are actions, are countable, is paralleled by the corresponding difficulty about material objects or things, which may be uncountable, while pies and bicycles and men, which *are* material objects, may be perfectly countable. Both difficulties are spurious. Being countable or uncountable is, as Frege would tell us, not a property of objects, and there is not some curious character of 'being an act' or 'being an event' which justifies the erection of unheard of principles of individuation which would never be applied to sword-thrusts or dinners.

There is one noteworthy general point of method that has come up here. Given an argument about the individuation or identity of an event or action, we can often construct a parallel argument about the individuation of concrete things, which would be sound if the argument about events and actions were so, but which is patently unsound. Where an argument about events or actions can be tested thus it should be; we shall often be able to reveal latent nonsense in this way.

I now come to the one argument of Goldman which seems to me to be not merely frivolous. Indeed I suspect this argument may actually bring out why he wants to distinguish as two acts what he would surely not distinguish, as, say, two moves made in some historic chess game. It is his first argument: suppose someone, John, pulls a trigger, the gun goes off, the bullet hits Jim and he is killed. The act of pulling the trigger caused the event of the gun's going off. Also, John has killed Jim. If "John's act of killing Jim" is another description of the act also described as "John's act of pulling the trigger", then it ought to be just as possible to say "John's act of killing Jim caused the gun to go off" as "John's act of pulling the trigger caused the gun to go off". But it sounds quite bizarre. Therefore, Goldman concludes, "John's act of killing Jim" can't be another description of the same act as satisfies the description "John's act of (then) pulling the trigger".

It is true that the sentence sounds strange. However, it is obvious that whether a certain description is true of some event (or of some object) may depend on what happens at other times and places. It may sound bizarre to say that a man married his widow, or fought a duel with his widow's second husband, at a certain date. The reason is that we are using descriptions which

had not come true of the people in question at the time of the event being reported.

The same may be true of the description of an event. It is indeed not particularly odd to say "The widow stuck a knife into her husband", though it may amuse one to reflect that this might be paraphrased as "The thereby widowed lady stuck a knife into her husband". This would be precisely parallel to "The thereby act of killing Jim caused the gun to go off". If we say that John's act of killing Jim caused the gun to go off, the description in the subject-position is 'proleptic', applied by anticipation in view of what happened later: only when Jim was dead (or sure to die[8]) did the description "act of killing Jim" come true of the act of pulling the trigger; and, of course, it came true through the train of cause and effect. That is the feature that is not present in "He married his widow". It is however present in "The widow stuck her husband with a knife" – when it was *because* she stuck her husband with a knife that he died and she became a widow. Similarly it was because John caused the gun to go off that Jim died, and this somehow makes the proleptic definite description of his act sound inappropriate and out of key. But there is no logical difficulty, and the odd sound is removed by the paraphrase: 'That act which (as things turned out) was the killing of Jim by John caused the gun to go off.'

(8) Judith Jarvis Thomson thinks[9] that some difficulties about time and tense prevent our identifying, say, Sirhan's shooting of Kennedy with his killing of Kennedy. If A shoots B at t and B dies of it at t′, between t and t′ it is true to say that A shot B but not that A killed B.[10] There is no difficulty in this for my account. If it is not true to say between t and t′ that A has killed B, that will be because B isn't dead till t′. So much is agreed. But it merely means that, although that act has occurred which, as things turn out, will prove to have been an act of killing, things have not yet turned out so.

Quite analogously a father can't truly be called a grandfather before the birth of any grandchildren of his. Thomson herself makes this comparison: "something further has to happen, the children have to have children", and "something further has to happen after the shooting in order for the killing to have taken place – B has to die". The only difference between us here seems to be that I would say "in order for the shooting to have been an act of killing", rather than "in order for the killing to have taken place".

This difference is not quite trivial. For Thomson, there is something other than the act of shooting, the hit, the wound, the death, or anything in that line: it is *the act of killing*, which is an act that had not happened until B was

[8] If someone is not dead yet but sure to die as a result of what I have done, someone may say to me "You've killed him".

[9] 'The Time of a Killing', *Journal of Philosophy*, 58 (1971).

[10] Except (as in the Sirhan case) in what Thomson calls "the Hollywood use". This use is more extensive than that designation suggests: consider how we say "You've ruined the plan", "He's done for himself", "They'd blown it" – all of which relate to expected developments. But the point is not important because there are plenty of cases where it does not apply. Nothing here hangs on it.

dead. This does seem to be a case for applying Ockham's razor, lest philosophy become a pseudo-science discovering imaginary entities. It certainly leads Thomson into a problem which, after some labour, she recognizes as senseless. Or, as she puts it, there's no true answer to the question: *just* when did this act happen? She also gets involved in problems of continuity which, though interesting in themselves, have little to do with action-descriptions.

In another argument she refers to passives: The death came after the shooting; so, if the shooting was the killing, B died after A killed him – and so he died after he was killed! There is nothing in this. Even allowing the transition from "B died after that act of A's which (as things turned out) was an act of killing him" to "B died after A killed him", the change to the passive in this sentence does not generate an equivalent to it. *Pace* Noam Chomsky, passivization often does not produce an equivalent sentence.[11] For example, "In New York, A was informing B that p" isn't equivalent to "In New York, B was being informed by A that p". Thus the absurdity of a passivization is not generally a proof of the absurdity of the original sentence. Nor is it in the particular case we are considering. For of course a man dies *after* the performance of any act that could ever be called an act of killing him; and equally he can't be called "killed" *until* he is dead. These matters are a source of perplexity to Thomson, but raise no difficulty on a proper understanding of the many descriptions of an action, which depend on a variety of circumstances and other happenings – just as do the many descriptions of the same object. Thomson indeed seems to be in some difficulty even about these, from her strange language of "the time of realization of the father of quadruplets", and "the time of realization of the man who shot B", and from her speaking of A's *becoming* the man who shot B; after all, he was that man all along! What he became was: a shooter of B; also, it came about that "the man who shot B" was true of him.

Only of a (then) living man do we say he became a father or grandfather; but it can *come about* that someone was a father or grandfather, even though he is dead. This is a harmless form of change of the past: a mere 'Cambridge change': the difference in truth-value between "he was a grandfather" said in 1970 or "as of 1970, he was a grandfather", and "he was a grandfather" said in 1975, or "as of 1975, he was a grandfather", where it is the same long dead man that is being spoken of. And similarly, though an act is over, many things come to have been true of it, or there are many things it comes to have been as further happenings unfold.

Thomson's final argument is that "A's shooting of B is A's killing of B" requires that we regard "the time of completion of x's [verb]ing of y" as a non-extensional expression. – If "completion" means "consummation" it *is* non-extensional. For then it would mean "the time by which it had come

[11] This is now recognized among linguists where quantifiers are involved. For example, "Few boys like few girls" is not equivalent to "Few girls are liked by few boys".

about that [a given act of x] was a [verb]ing of y",[12] and, since the time would vary with the verb, the original expression "the time of completion of x's [verb]ing of y" will be non-extensional. That is not a 'heavy price to pay' but on the contrary an obvious truth. And the extensionality disappears when we expand our statement according to the explanation just given. If, however, we refer to the time of completion of an act, we usually mean the time by which the agent had completed his activity in the matter. Under that interpretation, the expression is extensional. For it is agreed that the agent doesn't have to engage in anything further; he has completed his part when he has pulled the trigger, and the consummation of the act under the description "killing B" is left to circumstances, or, as Davidson puts it, to nature.

In conclusion: I may seem to have let myself off too easy, not giving any account of the 'individuation' of actions or events. But it is not possible to do that, if it means fixing criteria for what is a single action or event. This is a natural consequence of the uncountability which is characteristic of the concept of action or event. On the other hand, suppose we take a countable concept of an action or event like, say, administering poison. Such an event will split up into a lot of sub-events or sub-actions; there might even be a gap in the process, which yet counts as one administration of poison – the administration being interrupted, say, by a fit of coughing on the part of the administerer. However, we are willing to count the whole episode as just *one* time that person administered poison, one administration. So here we have one action, and if what happens in consequence of it without any further contribution from the administerer yields reports using active verbs, "poisoned NN", "killed NN", which are true of him, the latter at least will yield a definite description of an action on his part, which is satisfied by that one original act of administering poison. There is here no promise of a theory of what is absolutely one and what are many actions and events; rather it appears that there is no such thing.

Nevertheless, even if a concept F isn't associated with a rule for counting F's, it is often still possible both to speak of the same F and even to attach numeral adjectives to the term, as when we say "But those were three distinct episodes!" So it might still be right to ask for a criterion of identity of F's, in spite of F's being an 'uncountable' concept. Davidson's offer: "events are the same if they have the same causes and effects" was made with warnings that he thought not much was achieved by it, just as not much is achieved by talk of 'spatio-temporal continuity' as a criterion of identity for *material objects*. I think that even so he over-estimates what is achieved in both cases. Since objects can travel and alter, a path from any region to any region which doesn't go through a vacuum will give us the identity at times t and t′ of occupants at either end of it. We may of course prissily add the restriction: so long as the travel involved is no faster than the speed of light, not being bold

[12] There would be nothing wrong with giving the act in the same way in the two places, so that we speak of 'the time which it had come about that x's φ-ing of y was a φ-ing of y'.

enough to use the identity and the travel involved to refute current physics. Spatio-temporal continuity is satisfied by such 'material objects'; it is therefore worthless as a criterion. Davidson's criterion of identity for events isn't a criterion in the sense of a 'way of telling'. It is rather a condition which must be met by identical events. Or, one would suppose, identical anything. Therefore what we are to glean from the 'criterion' is that if events x and y stand in the equivalence relation 'having the same causes and effects', then everything that is true of x is true of y and vice versa. But note that this would not serve to resolve a disagreement between Davidson and someone like, say, Goldman; Davidson will say the condition is met by John's speaking and his speaking loudly (for they are the same event) and the other will deny it.

I maintain that the demand for a criterion of identity of particular occurrences *just as such* is not a reasonable one. This can, I think, be seen in the following way. Suppose I say "There is a child magician". That is as much as to say that someone is a child and the same one is a magician. Quite generally we can read "For some x, φx and ψx" as "Something is φ and the same thing is ψ". But if you ask for a criterion of identity *here*, at any rate you won't be able to offer one in terms of the same predicates holding. Further, the use of "the same" here does not imply that we can give any definite description of an object which verifies our statement. (There may be several child magicians.) It is the same with events. The crucial question is, e.g., "Can one *predicate* 'being a blow' of a certain movement of a fist?" If so, and if one speaks truly, something was a movement of a fist and the same thing (event) was a blow. If we *can* construct a definite description of the fist movement and the blow being reported (if a particular one was being reported) then we can say that the fist movement and the blow satisfy these two definite descriptions and so also that here are two definite descriptions of one and the same event. But the use of 'same event' that I made in commenting on the *predication* stands independent of this business of constructing an identity statement.

Thus enemies of what I have said about these identities will then have to deny the predications which lie behind the identities. These considerations do allow me to give an account which has a pattern suiting *a large number* of cases where two descriptions are descriptions of the same event, and these are indeed exactly the ones which interest people. Did the subject make only one fist movement of a certain kind on an identifiable occasion? or did he strike only one blow then? Was his fist movement a blow? or, if he struck only one blow, was that a blow with his fist? If so, we can construct straightaway either the description "the fist movement which . . ." or "the blow which . . ." Say we can't do both immediately. Then we can take the one we can construct, and use it to construct the other.

Jaegwon Kim's enquiries, like Alvin Goldman's, seem to me to have a different target; I hope it may not be an illusory one. There is a use of the term "event" in talking about probability[13] in which one would for example

[13] But the trouble is, writers on probability have the most damnably various terminology.

distinguish between an event (or outcome: come to think of it, etymologically speaking event = outcome): getting a 6 and a 5 in a toss of two dice, and the event of getting a 6 on die A and a 5 on die B. Now let us suppose that I toss, and there are bets on both these different outcomes. The toss comes up 6 on die A and 5 on die B. This was just one toss and just one fall of the pair of dice on this occasion – but the event (in my sense) is an exemplification of those two different events (in the probability theorist's sense). It ought to be obvious that I have no quarrel with this way of talking about probability. But now, what does Kim do? He wishes, apparently, to speak of single events as particular occurrences, such, for example, that the same event can't (as a matter of grammar) happen twice. What can happen several times, I think he calls a generic event. If I understand him, then, he seeks to give an account of the identity of events that are 'particulars' such that the exemplification of two different 'generic events' is *eo ipso* two different events. Naturally a vast proliferation of Kimian particulars is achieved by this terminology. He need not regard it as a criticism of his position to say so (though such a criticism has been made). But all we have here is the point that there is a large number of events (in the probability theorist's way of talking) that any particular happening is. This is harmless and obvious, though perhaps it only has a useful application where your field of possibilities is nicely cut and dried. But if this is all, if I have hit the nail on the head, I fear it is pursuing a mere Will-o'-the-wisp to try and make out the account in connection with, say, someone's *having strolled* and his *having strolled in a leisurely fashion*. The material needs cutting and drying in order for us to get any worthwhile events in the probability theorist's style and, to my mind, one should not regard its moist and uncut state as posing metaphysical problems. Further: will not Kim himself want to be able to express what I express by saying that such-and-such different outcomes were exemplified together in a particular occurrence? He ought to see, then, that he has no disagreement with me but is engaged in another enterprise.

It will be seen that, while I am in agreement with Davidson that there are many descriptions of an action, we part company when it comes to his 'theory of event-identity'. Or again, his theory of adverbial modification. This really doesn't go at all well with the idea of many descriptions. For the adverbial modification that suits one verb may not consort well with another, and yet the two verbs may occur in different descriptions of the same action. Then you can't really break the connection between the adverbial phrase and the verb. Or a phrase may consort very well with both verbs and it may have different meanings – say different things – when you tie it to the one and to the other. Thus consider how A might have made a purchase from B *with cash*, and how this action might have been a bribing of C. (C, we will suppose, is B's brother and a judge.) Then, adopting Davidson's style: there is an x and a y such that x is a purchasing of y from B, and x is a bribing of C, and A does x and x is with cash. The "with cash" belongs too closely with the purchasing to do that. You can say if you like that the bribing, as well

as the purchasing, was an act done using cash, but that's not what "with cash" means in "purchase with cash" and "bribe with cash". In our story A did *not* bribe C with cash. Nor can one say: well, the account of adverbial modification doesn't work *everywhere*, but it is of some value wherever it does work. For it may work merely because there are not currently two verbs in the language, with which an adverb associates with different effects. But one cannot be sure that there never will be two such verbs, or lay it down that there *shall* not be. The only adverbial phrases which are clearly immune to this possibility are specifications of place and time. For these we may gladly accept Davidson's treatment.

Thus I too treat events and actions as individuals, and there is agreement about the many descriptions of the same. But it seems impossible to accept the main further features of Davidson's logical theory of action sentences; we should reject his theory of event identity (none such being called for), while it seems to be demonstrably wrong in principle to separate adverbial phrases from their verbs in most cases.[14]

[14] I have an indebtedness to Kripke in connection with the final point; indirect in two ways. One, that I profited from a conversation with Robert Hambourger, who told me an example of Kripke's, and two, that Kripke's example (which I will not reproduce) constructs an absurdity for Davidson, arising precisely out of the multiple description thesis.

20 *Analysis* Competition – Tenth Problem

"It is impossible to be told anyone's name. For if I am told 'That man's name is "Smith"', his name is mentioned, not used, and I hear the name of his name but not his name."

REPORT

There were fourteen answers sent in; one is publishable, but is not a solution. Its emendation of the problem should certainly be adopted.

Entries were disqualified for publication by arguing against the conclusion, which everyone knows to be false in any case; for arguing that there is some truth in the conclusion, for the same reason; for displaying ignorance of the necessity and importance of the customary distinction of use from mention; for a solution based on the fact that we do not hear quotation marks (these are the sign, not the substance, of the difference of reference which makes logicians speak of the word's being a different name); for suggesting that we ought to be able to say that we hear a name if we hear its name (for the Jews, something is the name of God's name, which remains hidden); for introducing the point that I may learn a name by hearing it used (irrelevant unless hearing Smith called by his name is supposed to tell me what name the name of his name is a name of; but suppose I hear him called by a nick-name and then am told "His name is 'Smith'", am I to infer that the name named in that piece of information is the nick-name I have heard used?); for saying that it is only because names of words are names of *words* that an oddity seems to arise – i.e. the difference between (our customary) names of words and names of other objects is illusory; for saying that the argument is "too good" because it is of such general application; for excessively vague remarks about "situations" and so on; for talking about how "Smith" is the name of Smith and leaving the question how his name can be spoken of untreated or only vaguely referred to; and generally for fallacy in regard to the particular point which the puzzle is calculated to elicit.

This point was thoroughly stated by K. Reach[1] in 1938. Reach gives a name-table, remarking that "instead of 'table' one should say 'museum', for a table contains an arrangement of names, whereas a museum places things and names together". The table consists of two rows. In the upper row there is a representative of each 'figure' (i.e. single symbol) of a language;

[1] 'The Name Relation and the Logical Antinomies', *Journal of Symbolic Logic* (September 1938). I am indebted to the Editor for permitting the quotation from this article.

From *Analysis*, 16, 6 (1956) and 17, 3 (1957).

beneath is a word which is a representative of the 'figure' which is a name of the 'figure' represented above it. I give a section of Reach's table, enough to reproduce that part of his paper which has to do with the competition problem.

;	Semicolon	Secol
Semicolon	Secol	Sco

The SZ[2] expresses the relation between neighbouring figures above the line, by joining the corresponding names under the line with the words "is the name of".

Defects of the SZ. The purpose of the sentence "Secol is the name of Semicolon" is to give information about the meaning of Secol (i.e. "Semicolon"). Does this sentence serve its purpose? Suppose somebody asks "What is the meaning of Secol?" and he receives the answer "Secol is the name of Semicolon". If the answer is to convey anything to the questioner, it must be understood, i.e. the questioner must know what Sco and Secol stand for in the sentence. That he knows the former is shown by the *form* of his question, but the *meaning* of his question is that he does not know the latter. Hence the answer is incomprehensible to the questioner.

This could be objected to on the ground that it is not the business of language to explain the meaning of its own symbols, and that whoever makes use of a language must understand its words. The reply to this is that a language has to perform the tasks that are asked of it, and, further, if it is not the business of a language to explain the meaning of its symbols then the introduction of the SZ would be useless.

There is a single (very slight) contribution to the subject in Wittgenstein (*Philosophical Investigations*, 16):

> What about the colour samples that A shows to B: are they part of the *language?* Well, it's as you please. They do not belong among the words; yet when I say to someone: "Pronounce the word 'the'" you will count the second "the" as part of the sentence. But it has a role just like that of a colour-sample in language-game (8); that is, it is a sample of what the other is meant to say.

But the solution indicated here is only a clear one if its application is very restricted, as it is in this passage. It is not clear how it could be applied to, e.g. "'Red' is the name of a colour". Certain contributors looked in this direction.

The winning entry is by Al. Tajtelbaum, of New York. This entry just avoids disqualification under the last head by mentioning that our convention is tacit. I am not sure how aware this author is of the point made by Reach; he says that in formal analysis we don't ask how we know what name is named by the name of Smith's name, which seems to imply that elsewhere we can; in his positive contribution (third paragraph) he neatly avoids considering names of names of names. Thus he avoids considering the statement

[2] Carnap's *syntaktische Zuordnung*.

that "'Smith'" is the name of 'Smith' to which statement Reach's argument relates; but this is a particular case of the generalization that a name and its name are denoted by the same word – our tacit convention. It appears that the convention not merely is but must be tacit: a case of "What can be shown cannot be [informatively] said".

The demonstration that our (tacit) convention is not essential is correct; but it is equally clear that in general the names of names could not usefully be like the Jewish name of God's name (let the name of God's name be "S", and consider the question "What is S?" with the answer "God's name"); there must be a systematic connection between a name and its name such that a person can form the name of the name from mere acquaintance with *it*, and know what name the name of a name is a name of on hearing it. This contrasts with what ordinarily holds for objects and their names.

Mr Tajtelbaum suggests that we may take it for granted that an association between words and what they stand for has been established. This may be all right. But if 'association' is regarded as a uniform concept, when in fact it covers things whose variety is of logical importance, then it is possible that actual difficulties may arise simply from a failure to enquire into differences. I do not know if Mr Tajtelbaum is justified in speaking of a limitation of formal analysis here; whether he means that the practitioners of it merely *do* not deal with any of these questions, or that it can be seen straight off that they *could* not usefully try to do so, at least in some cases (such as this one). Reach's work seems to be a case of such an attempt.

"IT IS IMPOSSIBLE TO BE TOLD ANYONE'S NAME"

Al Tajtelbaum

Miss Anscombe's "problem" expresses a paradox from the following point of view. On the one hand, at least the cruder elements of society convey Smith's name by drawing attention to him (pointing, staring) and actually using Miss A.'s words: "That man's name is 'Smith'". On the other hand, according to the customary distinction between names and names of names, we only hear the name of Smith's name.[3] So, how do we learn his name? Alternatively, since in fact we do learn his name in this way,[4] how can the customary distinction be maintained?

The answer is plain: we have the (tacit) convention that a name and its name are denoted by the same word, and so the name of a name "tells" us the name.

[3] To see this observe that the sense of "That man's name is 'Smith'" is altered if we substitute one of Smith's other names for "Smith".

[4] The unqualified "impossible" in the first sentence of Miss Anscombe's problem is a little vague; it seems best to add to this sentence "in the following way", and drop "For" from the next.

The question arises whether this convention is essential. The following trivial example shows that it is not. Observe first that Miss Anscombe's problem makes sense if we restrict ourselves to words which are either (1) not names at all, (2) names, (3) names of names, that is we do not consider names of names of names. Also, we may restrict outselves to spoken language. Suppose now that Smith's name is pronounced as a single sound in the usual way while the name of his name is always spelled out letter-by-letter: S-M-I-T-H. Then, in order to learn Smith's name from "That man's name is S-M-I-T-H" it is sufficient to have a preliminary training in stepping from a letter-by-letter presentation of a word to the ordinary (continuous) reading of the word.

It goes without saying that in a formal (logical) analysis of language one ignores the question how words (more precisely, those words which refer to things) are associated with the things to which they refer; but takes for granted that such an association has been established. One neither asks how we learn or know who is named by the words "that man" nor how we learn or know what (name) is named by the name of Smith's name. One does not even examine what exactly is meant by saying that such an association has been established. It is possible that Miss Anscombe is more interested in the nature of this association than in the formal analysis of a "given" language: if so, her problem has an additional interest for her and those who share her taste, because it draws attention to a limitation of formal analysis.

New York

21 A Reply to Mr C. S. Lewis's Argument[1] that "Naturalism" is Self-Refuting[2]

I want to discuss your argument that what you call "naturalism" is self-refuting because it is inconsistent with a belief in the validity of reason. With this argument you propose to destroy 'naturalism' and hence remove the determinist objection to miracles.

For my purpose it is not necessary to go into your description of 'naturalism' or your claim that one must either believe it or be a 'Supernaturalist' – i.e. believe in God. For you say that 'naturalism' includes the idea that human thought can be fully explained as the product of natural (i.e. non-rational) causes, and it is this idea which you maintain is self-contradictory because it impugns the validity of reason, and therefore necessarily of any thinking by which it itself is reached.

What I shall discuss, therefore, is this argument: the hypothesis that human thought can be fully explained as the product of non-rational causes is inconsistent with a belief in the validity of reason.

You state it as a rule that "no thought is valid if it can be fully explained as the result of irrational causes", and you give examples to show that we all universally apply this rule: we do not attend to the belief of a man with delirium tremens that the house is full of rats and snakes; we are less impressed by a man's gloomy views if we know he is suffering from a bad liver attack; the disruptive power of Marxism and Freudianism against traditional beliefs has lain in their claim to expose irrational causes for them.

About the first two examples I should like to say that it is only because we already know that men with delirium tremens see things that are not there, and that men with liver attacks take gloomier views of the situation than they would normally take, or than is reasonable, that we dismiss a man's belief by ascribing it to delirium tremens or to a liver attack when we know he has one of these complaints.

What sorts of thing would one normally call "irrational causes" for human thoughts? If one is asked this, one immediately thinks of such things as passion, self-interest, wishing only to see the agreeable or disagreeable,

[1] A short version of Lewis' argument can be found in his paper "Religion without Dogma" on pp. 87–8 of this issue of the *Socratic Digest* which is also commented on by Prof. H. H. Price in his reply to that paper, section 4, p. 98. The argument in full is in "Miracles", Chapter III of the first edition.

[2] I wish to acknowledge that I was very greatly helped in writing this paper by discussing it with Mr Y. Smythies: naturally, he is not responsible for its faults.

Socratic Digest (Oxford) [*c.* 1947].

obstinate and prejudicial adherence to the views of a party or school with which one is connected, and so on. Suppose one mentions such things, and then someone says: "There are also tumours on the brain, tuberculosis, jaundice, arthritis and similar things", one would rightly object that these do not belong in the same list as the others. They are not "irrational causes", they are conditions which we know to go with irrational beliefs or attitudes with sufficient regularity for us to call them their causes.

You speak of "irrational causes", and by that you seem to mean "any cause that is not something rational". "Something rational" you explain by example: "such as [you say] argument from observed facts". You contrast the following sentences: (1) "He thinks that dog dangerous because he has often seen it muzzled and he has noticed that messengers always try to avoid going to that house;" (2) "He thinks that dog dangerous because it is black and ever since he was bitten by a black dog in childhood he has always been afraid of black dogs."

> Both sentences [you say] explain *why* the man thinks as he does. But the one explanation substantiates the value of his thought and the other discredits it. The difference is that in the first instance the man's belief is caused by something rational (by argument from observed facts) while in the other it is caused by something irrational (association of ideas).

I am going to argue that your whole thesis is only specious because of the ambiguity of the words "why", "because" and "explanation". That ambiguity is illustrated here. The case of the man who is frightened by black dogs is unclear. Imagine the two following possibilities: (1) He says "That dog's dangerous." He is asked, "How do you know?" He says, "It's black: I was once bitten by a black dog." To this we reply: "That's not a good ground. We know enough about dogs to know that." (2) He says: "That dog's dangerous." He is asked, "How do you know?" But to this he gives no answer; he shakes his head, trembles and says, "It's dangerous." Then either he, or someone else, says that he behaves like this because he was once bitten by a black dog. Then we can know that we need not pay attention to his belief; it already appeared groundless, from the fact that he could give no grounds; but now we are satisfied it is groundless because we understand it as the expression of a fear produced by circumstances which we know to give no good grounds for fear. It is here quite natural to speak of "irrational causes".

Similarly it is true that the Marxists and the Freudians claim to expose irrational causes for various traditional beliefs. The Freudian says that my belief in God is a projection of my infantile attitude towards my father. The Marxist says that many of my beliefs and reasonings arise from my considering things important that I should not consider important if I were not bourgeois, and neglecting other things which I should not neglect if I were not bourgeois; and that the whole point and significance of certain kinds of thinking is simply – by the very pretence of detachment that they make – to draw people away from relating their thoughts to the class-struggle.

But by your equation of "irrational cause" with "non-rational cause", you are led to imagine that if the naturalist hypothesis (that all human behaviour, including thought, could be accounted for by scientific causal laws) were true, human thought would all have been explained away as invalid; that if human beings could be shown to act according to such laws, their case would have been shown to be universally like the particular case of the man who is actuated by "irrational causes" and whose beliefs are groundless. This seems to me to be a mistake founded on various confusions you commit about the concepts of "reason", "cause" and "explanation"; and I hope by showing what the confusions are, to show that it *is* a mistake.

First, I want to examine your remark that we must believe in the validity of reason, and that we can see when a hypothesis is inconsistent with a belief in the validity of reason, and refute it by the consideration that it is inconsistent with that belief.

You can talk about the validity of a *piece* of reasoning, and sometimes about the validity of a *kind* of reasoning; but if you say you believe in the validity of reasoning itself, what do you mean? Isn't this question about the validity of reasoning a question about the validity of *valid* reasoning? Suppose that you are asked to explain "valid", how will you do it? The most obvious way would be to show examples of valid and invalid reasoning, to make the objections which, in the examples of invalid reasoning, show that the conclusion does not follow from the premisses; in the cases of valid reasoning, to elucidate the form of the argument: if the piece of reasoning under consideration is elliptical, to add the statements which are required to enforce the conclusion. Whether you would adopt this method or some other (though I do not know of any other), I suppose you think it *somehow* possible to explain to yourself or someone else what "valid" means, what the distinction between "valid" and "invalid" is? Now if the naturalistic hypothesis (that human thought is the product of a chain of natural causes) is proposed to you, you say: "But if this were so, it would destroy the distinction between valid and invalid reasoning." But how? Would it imply that you could no longer give the explanation you gave, point to and explain the examples, say which arguments proposed to you are valid and which invalid in just the same way as you did before the naturalistic hypothesis was supposed? "But," you may say, "though I should of course know which arguments to *call* valid, or which I should have *called* valid, I should not now feel any confidence that they were *really* valid." But what do you mean by "really valid"? What meaning of "valid" has been taken away from you by the naturalistic hypothesis? What *can* you mean by "valid" beyond what would be indicated by the explanation you would give for distinguishing between valid and invalid, and what in the naturalistic hypothesis prevents that explanation from being given and from meaning what it does?

You say that on this hypothesis there would be no difference between the conclusions of the finest scientific reasoning and the thoughts a man has because a bit of bone is pressing on his brain. In one way, this is true.

Suppose that the kind of account which the "naturalist" imagines, were actually given in the two cases. We should have two accounts of processes in the human organism. "Valid", "true", "false" would not come into either of the accounts. That shows, you say, that the conclusions of the scientist would be just as irrational as those of the other man. But that does not follow at all. Whether his conclusions are rational or irrational is settled by considering the chain of reasoning that he gives and whether his conclusions follow from it. When we are giving a causal account of this thought, e.g. an account of the physiological processes which issue in the utterance of his reasoning, we are not considering his utterances from the point of view of evidence, reasoning, valid argument, truth, at all; we are considering them merely as events. Just *because* that is how we are considering them, our description has in itself no bearing on the question of "valid", "invalid", "rational", "irrational", and so on.

Given the scientific explanation of human thought and action which the naturalist hypothesis asserts to be possible, we could, if we had the data that the explanation required, predict what any man was going to say, and what conclusions he was going to form. That would not mean that there was no sense in calling what he did say true or false, rational or irrational.

> But [you say] this imagined explanation would show that what we said was not caused by reason but by non-rational processes. We may give arguments, but, as everything we say will be fully explained by non-rational causes, (1) the idea that conclusions are derived from premisses will be an illusion (hence I say that the explanation impugns the validity of reason) and (2) the idea that we think what we do because of reasoning, i.e. because we have reasoned, will be an illusion. Every thought will have been produced by a non-rational chain of causes and therefore not by such rational causes as observation and argument. So no thought will be worth anything.

I want to say that such an argument as this is based on a confusion between the concepts of cause and reason, which arises because of the ambiguity of such expressions as "because" and "explanation".

(1) If I said: "You think this conclusion follows from these premisses, but in fact the assertion of it is a physical event with phsyical causes just like any other physical event," would it not be clear that I was imagining the ground of a conclusion to be a kind of cause of it? Otherwise there would be no incompatibility: "this conclusion follows from these premisses" would be in no way contradicted by "the assertion of this conclusion is a physical event with physical causes like any other physical event". Even though all human activity, including the production of opinions and arguments, were explained naturalistically, that could have no bearing on 'the validity of reason' – i.e. on the question whether a piece of reasoning were valid or not. Here I am speaking of 'reason' in a non-psychological sense, in which 'a reason' is what proves a conclusion. If we have before us a piece of writing which argues for an opinion, we can discuss the question: "Is this good

reasoning?" without concerning ourselves with the circumstances of its production at all.

(2) But you may say that you do not wish to call a reason – in this non-psychological sense – the cause of its conclusion; you may agree that the naturalistic hypothesis could not impugn 'the validity of reason' in this sense, but say that it makes reason an ideal which we cannot attain; that it does impugn the validity of all *actual* human reasonings. For granted that the logical derivation of a conclusion from its premisses could not be affected by any hypothesis, yet if *our reasonings* are to be valid *we* must derive the conclusions from the premisses, in actual fact.[3] This introduces a psychological application of the concept of "reason" which is used if we ask the following questions about, e.g., a piece of writing that we are examining: "Granted that this is a piece of reasoning, did the man who wrote it actually *reason?* Was he really persuaded by this reasoning or by something else? Or – another possibility – did he really understand and mean this argument? Or did he perhaps write it down quite mechanically? Here is a statement (which is even in fact correct), but did the man who wrote it himself assert it because of the good grounds which do exist for asserting it?" If we can answer "yes" to such questions as these we call the opinions in question "rational" or the man "rational" for holding them. And if we know that a man's opinions are not rational in this sense, we regard it as accidental if in fact they are worth attending to or true; we shall not expect to find them worth attending to, and if it is a question of information to be accepted on his word, we shall not accept it.

You argue that the naturalist hypothesis about human thinking implies that no human thinking is rational in this sense. For if a man produces what purports to be the conclusion of an argument, in order that what he says should be rational he must say it *because* he has reasoned; but the naturalist hypothesis says that he says it *because* of certain natural causes; and if these causes *fully* explain his utterance, if the chain of causes is complete, there is no room for the operation of such a cause as the man's own reasoning. So someone might say: "If I claimed to be able to kill a man by an act of will, and he died, but his death was fully explained by the fact that someone who had sworn to murder him shot him through the heart, that would demolish my claim to have killed him by an act of will."

Your idea appears to be that '*the* explanation' is everywhere the same one definite requirement: as if there were a fixed place for '*the* explanation' so that we can know, when it is filled, that, if it has been correctly filled, the whole subject of 'explaining this fact' has been closed. We understand the requirement antecedently to any knowledge of the kind of investigation that might be made, and, once we see that the requirement has been satisfied, no further question can be asked.

But the concept of 'explanation' has very varied applications, and the expression "full explanation" has reference only to the type of explanation

[3] I think this is the argument of Mr Lewis' reply to me.

that is in question. I may, for example, ask a man to explain to me his reasons for thinking something. He gives me an explanation. I may say: "That's not a full explanation; there must be more to it than that – for it explains, let us say, why you take a naturalistic view but not why your view is a phsyical or physiological naturalism; the arguments you have given are consistent with a psychological naturalism: tell me why you reject that." Now if I ask for this sort of investigation I am not making a *causal* enquiry at all: I am asking for grounds, not causes; and you can only have imagined that it was appropriate to speak of "causes" because the word "because" is used. Giving one's reasons for thinking something is like giving one's motives for doing something. You might ask me: "Why did you half-turn towards the door?" and I explain that I thought I saw a friend coming in, and then realized it was someone else. This may be the explanation although I did not at the time *say* to myself "Hello! There's so-and-so; I'll go and speak to him; oh no, it's someone else." So when I give the explanation it is not by way of observing two events and the causal relation between them.

The naturalistic hypothesis is that causal laws could be discovered which could be successfully applied to all human behaviour, including thought. If such laws were discovered they would not show that a man's reasons were not his reasons; for a man who is explaining his reasons is not giving a causal account at all. "Causes", in the scientific sense in which this word is used when we speak of causal laws, is to be explained in terms of observed regularities: but the declaration of one's reasons or motives is not founded on observation of regularities. 'Reasons' and 'motives' are what is *elicited* from someone whom we ask to explain himself. Of course we may doubt that a man has told, or even made clear to himself, his real reasons and motives; but what we are asking for if we say so is a more searching consideration, not an investigation into such a question as: "Is this really an instance of the causal law which I have applied to it?" – and that is true even though, as is possible, we doubt him on grounds of empirical generalizations which we have made about people's motives and reasons for the action or opinion in question. Such generalizations are possible, and hence one can imagine a psychological naturalism which believes in the possibility of a complete scientific system of psychological causal laws of human behaviour. It is important to realize that such a notion of psychological causality (which would arise from observing regularities in people's motives and mental processes) should be distinguished from the use of "because" in the *expression* of motives and mental processes.

It appears to me that if a man has reasons, and they are good reasons, and they are genuinely his reasons, for thinking something – then his thought is rational, whatever causal statements we make about him. Even though he give good reasons, however, we may detect in him such passions or such motives of self-interest in saying what he does that we say that it is not really "for these reasons" that he says it, and regard the reasons as a façade that he puts up to obscure his 'real reasons': these 'real reasons' being the kind of

thing that I admitted as 'irrational causes'. And we rightly suspect and scrutinize carefully the reasoning that he offers. Or we may think him so dominated by 'irrational causes' that it is not worthwhile to look at his reasoning at all: though the mere fact that he is actuated purely by these motives does not necessarily mean that he will not in fact be able to reason well.

So far I have only talked of a man's reasons in a sense in which: "He thinks so-and-so because of such-and-such a chain of reasoning" is in no way a causal statement. There is a kind of statement that I have not yet considered, which is in some sense causal. Suppose I ask someone why he believes something, and he begins to produce reasons, I may say: "Sorry, I didn't mean that – I know what reasons there are for believing as you do; what I meant to ask was what in actual fact, as a matter of history, led you to this opinion, what *caused* you to adopt it?" This is a quite intelligible question which anyone would know how to answer. It seems to me that you have not distinguished it, as it ought to be distinguished, from the question "What are your reasons?" – and that it is in virtue of his answer to the latter that a man or his opinions should be called rational, whatever his answer to the former. However, as your argument stands, it says that human thought is discredited unless his answer to the former question ("What, as a matter of history, led you to this belief?") states the *occurrence* of reasoning; and you also argue that on the naturalist hypothesis an answer which does state the occurrence of reasoning cannot be true, because the naturalist hypothesis is that non-rational causes produce his opinions. I should also deny this part of your argument. For though it is natural to use the word "cause" here, the logic of "cause" as used here is different from its logic as used when we speak of causal laws. Suppose someone asks me for such a historical account of the mental processes which actually issued in my belief, and I give it to him. And suppose he then asks: "What reason have you for calling the thing that you mention in answer to this question the *causes* of your belief?" At first I would imagine that he was accusing me of self-deception, saying, "Look into it more thoroughly and you will realize that you have not given a truthful account." But suppose he makes it clear that he is not suggesting anything of this kind; he does not doubt my account of my mental processes at all; but, given that they occurred just as I have related them, and that afterwards I held the opinion which I say resulted from them, he asks why I say that it did result from them, that they did produce it? Would this not be an extraordinarily odd question? It makes it seem as if one made here a causal statement analogous to scientific causal statement, which would be justified by – roughly – appeal to observed regularities; but here, though it is natural, given the kind of question "What actually led you to this?" to speak of a "cause", yet the sense of "cause" as used here is not to be explained by reference to observed regularities. That is sufficient to show that this is one more case of the great ambiguity of "explanation", "why", "because," and "cause" itself. And therefore the discovery of scientific causal laws could not

demonstrate the falsity of such assertions as "I thought so-and-so as a *result* of such-and-such consideration".

I do not think that there is sufficiently good reason for maintaining the "naturalist" hypothesis about human behaviour and thought. But someone who does maintain it cannot be refuted as you try to refute him, by saying that it is inconsistent to maintain it and to believe that human reasoning is valid and that human reasoning sometimes produces human opinion.

A causal explanation of a man's thought only reflects on its validity as an indication, if we know that opinions caused in that way are always or usually unreasonable.

DISCUSSION

In his reply Mr C. S. Lewis agreed that the words "cause" and "ground" were far from synonymous but said that the recognition of a ground could be the cause of assent, and that assent was only rational when such was its cause. He denied that such words as "recognition" and "perception" could be properly used of a mental act among whose causes the thing perceived or recognized was not one.

Miss Anscombe said that Mr Lewis had misunderstood her and thus the first part of the discussion was confined to the two speakers who attempted to clarify their positions and their differences. Miss Anscombe said that Mr Lewis was still not distinguishing between "having reasons" and "having reasoned" in the causal sense. Mr Lewis understood the speaker to be making a terachotomy thus: (1) logical reasons; (2) having reasons (i.e. psychological); (3) historical causes; (4) scientific causes or observed regularities. The main point in his reply was that an observed regularity was only the symptom of a cause, and not the cause itself, and in reply to an interruption by the Secretary he referred to his notion of cause as "magical". An open discussion followed, in which some members tried to show Miss Anscombe that there was a connection between ground and cause, while others contended against the President that the test for the validity of reason could never in any event be such a thing as the state of the blood stream. The President finally admitted that the word "valid" was an unfortunate one. From the discussion in general it appeared that Mr Lewis would have to turn his argument into a rigorous analytic one, if his notion of "validity" as the effect of causes were to stand the test of all the questions put to him.

NOTE BY C. S. LEWIS

I admit that *valid* was a bad word for what I meant; *veridical* (or *verific* or *veriferous*) would have been better. I also admit that the cause and effect relation between events and the ground and consequent relation between propositions are distinct. Since English uses the word *because* of both, let us

here use *Because CE* for the cause and effect relation ("This doll always falls on its feet *because CE* its feet are weighted") and *Because GC* for the ground and consequent relation ("A equals C *because GC* they both equal B"). But the sharper this distinction becomes the more my difficulty increases. If an argument is to be verific the conclusion must be related to the premisses as consequent to ground, i.e. the conclusion is there *because GC* certain other propositions are true. On the other hand, our thinking the conclusion is an event and must be related to previous events as effect to cause, i.e. this act of thinking must occur *because CE* previous events have occurred. It would seem, therefore, that we never think the conclusion *because GC* it is the consequent of its grounds but only *because CE* certain previous events have happened. If so, it does not seem that the *GC* sequence makes us more likely to think the true conclusion than not. And this is very much what I meant by the difficulty in Naturalism.

Index

act of acknowledgement, 99
action,
  intention and, 75–82
  "under the description", 208–19
admissibility, 96
adverbial modification, 218–19
"after", grammar of, 180–95
Altham, Dr J., 29n
ambiguity, 184–5
Analysis of Mind (Russell), 127n
Analytical Philosophy (*ed.* Butler), 3n
anger, pretence and, 87, 88, 88n, 89–90
antecedent, subjunctive conditionals, 197–9, 201, 204
appearance,
  false, 91
  subjectivity of sensation, 47
  substance and, 39–40
Aquinas, St Thomas, vii, 88n, 152–3
Aristotle, 46, 50, 87, 208
  causation, 134, 141
  substance, 38, 41
asymmetry, of verbs, 60–1
atomism, 151
Augustine, St, 21, 21n, 32
Austin, John, 17, 42
  pretending, 83, 83n, 84–7, 89–91
Ayer, A. J., 31n, 37

Basson, A. H., 38n
"because", naturalism and, ix, 225, 227, 229, 230–1
"before", grammar of, 180–95
beginning, 193–5
beginning of existence, 148–62
being, concept of, vii
belief,
  as event, 63
  grounds and, x
  memory and, 126, 127
Bennett, D., 209n
Bennett, Prof. Jack, x
Berkeley, George,
  'idea' of self, 55
  perception, 64–5
  sense-impressions, 11, 16
Bierce, Ambrose, 31n
body, mind and, 21–2, 31, 33–4, 57
Bogen, J., 209
Born, Max, 145
Braybrooke, D., 71, 71n, 72
Brentano, Ludwig, 99
Broad, C. D., 122, 122n, 123, 125
  determinism, 145
Brown, S. C., 64n
Butler, R. J., 3n

Carnap, R., 173n, 221n
Cartesian Ego Theory, 25, 31
Castaneda, H-N., 22n
causal condition, choice and, 166–7, 171
causal power, 163
causal statement, 178
causality, ix–x
  and determination, 133–47
  and extensionality, 173–9
  and naturalism, 229
causation,
  'experience' and, 120–30
cause, vii, 3; *see also* 'irrational causes'
  beginning of existence, 148–62
  'mental', 75–82
  naturalism, 226–7, 229–32
cause and effect relation, *see* effect
certainty, 53, 144
changelessness of past, 113, 116
choice, as causal condition, 166–72
Chomsky, Noam, 215
Church, Alonzo, 173n
Cioffi, F., 14
Clarke, Thompson, 18
cogitatio, 60–1, 121; *see also* thought
colour; *see also* 'standing colour of objects'

object of sight, 3, 14, 15–17
subjectivity of sensation, 44–7, 49, 52
substance and, 39, 42
coming into being, concept of, vii
communication, about inner event, 61
concept, cf. judgement, 99
Concept of Mind, The (Ryle), 79n, 83n, 143n
'conception', first person and, 29
conclusion, 228, 232
conditional, 163
sense of the, 111
subjunctive, 196–207
connection, predicative mode of, 94–9
consciousness, cf. self-consciousness, 36
consequent, subjunctive conditionals, 197–9, 201
Contemporary British Philosophy (Lewis), 120n
continuity, 215
continuum, Hume's rejection of, 151
contraposition, subjunctive conditionals, 204
count-noun, event as, 213
criterion,
'criterion within', 52–3
subjectivity of sensation, 51
Critique of Pure Reason (Kant), 135
cynicism, 93

D'Arcy, Fr Martin, viii
David Hume (Basson), 38n
Davidson, Donald, 147, 163, 163n, 164, 173, 174, 178
action, 211, 216–19
description; *see also* 'definite description', 'internal description'
Russell's theory of, 176
determinism, soft, 163–72
De Trinitate (Augustine), 21
'definite description', 95, 98
"I" as, 32
demonstrative, first person and, 27–8
depth, sense-impression, 16
Descartes, René, 3
cogitatio, 60–1, 63, 121
first person, 21–2, 25, 29, 31, 32, 35
substance, 37, 38, 41
determination, causality and, 133–47

distance, seeing and, 16, 65
"doubting", 21n
dualism, 21
duration, 158–9

Ebersole, Prof. Frank, 12n
effect, cause and, 80, 133, 134, 136, 138, 145, 148–53, 231–2
efficacy, 137–8
Einstein, Albert, 142, 144n
space, 154, 157
Elements of Philosophy Concerning Body (Hobbes), 134
ending, 193–5
"enjoy", grammar of, 94–100
equivalence, subjunctive conditionals, 202–6
Essays on Freedom of Action (*ed.* Honderich), 163n
essence, object and, 37
essentialism, 177–8
'ethical' freedom, 145–6
Ethics (Spinoza), 134
events,
and ground, 231
instantaneous, 191–3
in the mind, 57–63
and temporal clauses, 188–90
"under a description", 216–19
Evolution of Physics, The (Einstein and Infeld), 144n
expectation, 120
Russell, 64
experience,
memory and, 120–30
thought and, 58, 60–3
explanation,
C. S. Lewis, ix–x, 225–30
exportation, subjunctive conditionals, 204
expressions, predicative connection of, 94–100
extensionality, 216
causality and, 173–9
Eye and Brain (Gregory), 69

Feynman, R. P., 144, 145, 147
First Cause, existence of, vii

first person, 21–36
Föllesdal, Prof., 34n, 173
freedom, and determinism, 163n, 170–2
Freedom and Determinism (Lehrer), 163n
Frege, Gottlob, 7, 98, 213
first person, 23, 27n
Freudianism, 224, 225
friction, 148–50
From a Logical Point of View (Quine), 173, 211n
future, and reality of past, 112, 116

Geach, P. T., 9n, 175n, 182n
temporal quantifiers, 188n
Goethe, Johann Wolfgang von, 120, 121, 125–9
Goldman, Alvin, 163, 164, 211–13, 217
grammar of "enjoy", 94–100
grammatical understanding, 8
Gregory, Prof. R. L., 64–70
Gregory of Nyssa, St, 88n
Grice, H. P., 202
Griffin, J. P., 38n
ground, and event, 231, 232
Guttenplan, Samuel, 21n, 34n

Hacking, Ian, 135n
Haldane, J. B. S., ix
Hambourger, Robert, 219n
Hamilton, Sir William, 4
Harman, G., 7n
Havard, Dr, x
heat, friction and, 148–50
Honderich, Ted, 163n
Hobbes, Thomas, 61, 92–3
causality, 134, 153–4, 157, 159, 160
Hume, David, vii–viii, 32, 64, 129
causality, 76, 134–5, 137–9, 145, 148–54, 159, 159n, 160
idea of a thing's existence, 103
Hume's Theory of the External World (Price), viii
hypothesis, perception and, 65–6, 68–70

"I", *see* first person
idea, 3, 5, 9
Descartes, 60–1
sense perception and, 11

idealism, 128–9
identification, and subjectivity of sensation, 45
identity, viii
criterion of, 216–17
event-identity, 210–13, 218–19
first person, 22–3, 26, 33, 35
intentional object, 10–11
memory and, 128
perception and, 68
subjectivity of sensation, 46, 52–3
substance and, 38, 42
image,
language and, 62–3
veridical, 67–8
imagination,
and beginning of existence, 152–3, 160
and image, 63
as inhibited performance, 83–4
implicature, 202
impression,
sense-impression, 3–20
subjectivity of sensation, 44–6, 48–50, 52
substance, 43
inborn capacity, doctrine of, 97
'incorrigibility', of descriptions of impressions, 44, 46, 54
indeterminacy, 6
'indirect reflexive', 22, 29
individual, 'nominal essence' of, 38
individuality, substance and, 37, 42
individuation, of actions, 216
Infeld, Leopold, 144n
inference,
hypothesis and, 69
and sense-impression, 16, 54
input, perception and, 65–6, 68–9
Inquiry into Meaning and Truth, An (Russell), 64n
instantaneousness, 191–3, 195
intention, 57–9, 61, 63
language and, 209
reason and, 75–82
Intention (Anscombe), 71, 208n
intentionality, of sensation, 3–20
'internal description', 72, 73
introspection, first person, 34

'irrational causes', naturalism and, 224–6, 230

James, William, 12n, 36, 50, 71
John of Damascus, St, 88n
Joseph, H. W. B., 38
judgement, and subjectivity of sensation, 49–50, 51, 55, 56

Kant, Immanuel, 47
causality, 135, 145
representation of time, 103
void time, 158
Kehoe, Fr Richard, vii
Kenny, A. J. P., 143n, 208
Kim, Jaegwon, 217–18
knowledge,
memory as, 105–6, 109, 118, 127, 130
perception as, 120
'without observation', 71
Knowledge and Certainty (Malcolm), 130n
Kripke, Saul, 21, 21n, 29, 219n

language, 70, 221, 223
describing thought, 62
first person, 32
object-phrase and, 8
'ordinary language' doctrine, 13–15
Language, Truth and Logic (Ayer), 31n
Larsen, Sue, 96
laws of nature, causality and, 138–9, 146
Lehrer, Keith, 163, 163n, 164, 166–8, 171–2
Leibniz, 65, 208
Leonardo da Vinci, 43, 65
Lewis, C. S., ix–x, 224–32
Lewis, H. D., 120n
lie, 92
Locke, John, ix, 5
first person, 26
perception, 64–5
substance, 38, 43
Luckiesh, M., 12n

Malcolm, Norman, 130n
malleability, 40
Marxism, 224, 225
Mayo, Bernard, 203n
McTaggart, J. McT. E., 27n, 151
meaning,
of statements about past, 103, 105–6, 113, 117
subjectivity of sensation, 51
thought and, 58, 60
Meaning and Necessity (Carnap), 173n
mechanics, Newtonian, 139–40, 143, 144, 147
memory, 57–60
and 'experience', 120–30
and reality of past, 105–10
mention, cf. use, 220
Metaphysics (Aristotle), 134
Mill, J. S., 135
mind, events in, 57–63
Mind and its Place in Nature, The (Broad), 122n
Mind and Language (*ed.* Guttenplan), 21n
miracle, determinist objection to, 224
Miracles (Lewis), ix
Modal Logic (Hughes and Cresswell), 210n
Mondadori, Fabrizio, 196
Moore, G. E., 18, 163
motive,
backward-looking, 79–81
'mental cause' and, 76–81
Muller, Dr A., 69n
Mysticism and Logic (Russell), 135

name,
and name of name, 220–3
and proper name, 27, 32, 36, 54–5
naturalism, ix–x, 224–32
Nature of Belief, The (D'Arcy), viii
Nature of Existence, The (McTaggart), 27n
necessitation, cf. causation, 133–6, 139, 140, 144
negation,
external, 181
grammar of enjoy, 95, 96
of proposition, 187
Newton, Isaac, 159
non-substitutability, of different descriptions of object, 5

object,
intentional, 3–11

object—*cont.*
sense perception and, 11, 13, 17, 18, 64
substance and, 37
observation, and subjectivity of sensation, 54
Our Knowledge of the External World (Russell), 188n

pain, 89
as *cogitatio*, 60
Parmenides, 103
passivization, 215
past,
memory and, 123, 124, 125, 129
reality of, 103–19
Peirce, C. S., 135n
perception, viii, 14, 19, 46, 231
causality and, 137
enjoyment of substance, 99
Gregory and, 64–70
knowledge and, 120
memory and, 107
performance, pretending and, 83, 84, 93
personal pronoun, 23
phenomenalism, viii–ix
Aristotle, 50
Plato, 54, 120
Philosophical Investigations (Wittgenstein), 52, 76, 221
Philosophy of Psychology (*ed.* Brown), 64n
place, of something coming into existence, 154–7, 160–2
plane, concept of, 65
Plato, 15, 54, 80, 120
anger, 88n
sensory judgement, 50–1
position, sensation of, 71–4
possibility, Hume, 152–3
posteriority, *see*, "after"
predicate,
and grammar of "enjoy", 95, 98
substance-involving, 40–1
predication, 38
presence to consciousness, 28
pretending, 83–93
Price, H. H., viii, 19, 224n
Principia Mathematica (Russell), 176
Principles of Philosophy (Descartes), 21n
Principles of Psychology (James), 36n
Prior, Prof. Arthur, 9n
Prior Analysis (Aristotle), 208
priority, 183
'private ostensive definition', 52–3, 85
probability, 217, 217n
proper name, 38
first person and, 23–4, 26–7, 27n, 28–9, 55
'proper sensibles', Aristotle, 46, 50
proposition, 186–8
cf. concept, 99, 100
disjunction of, 205
effect presented in, 178–9
present-tense, 181–2

quantum theory, 142, 145
Quine, W. V. O., 16, 173, 176–7, 176n, 178
intensional objects, 211

Reach, K., 220–1
reality,
of past, 103–19
pretence and, 85
reason,
intention and, 75–6, 81–2
naturalism and, 224–32
validity of, 226–8, 231
recognition, 16, 231
reference,
first person, 28–9
'guaranteed', 29–31
referring expression, 31–2
reflexive pronoun, first person and, 22–3, 25–6
'rembering', 121–2, 124–6
repetition, 186, 187
revenge, 79–80
Rhees, R., 111
'rule of sequence of tenses', 86
Russell, Bertrand, viii, 31, 32, 176, 177, 177n
law of causation, 135
'mnemic causality', 127n
perceptive experience, 64, 64n
sense-data, 3, 16
substance, 37
time, 188, 190–1, 193

Ryle, Prof. Gilbert, 143n
achievement verb, 192
motive, 79
pretending, 83, 84

"same", 53, 104
Schopenhauer, Arthur, 164
Schrödinger, 142
Schwartz, Robert, 46
Science and Humanism (Schrödinger), 142
scientia media, doctrine of, vii
secondary qualities,
metaphysical status, 44, 46
'sensible', 39, 40, 41
'sensible', 39, 40, 41
seeing; *see also* sight
and intentionality of sensation, 12, 13, 15, 16–20
seeming, pretending and, 85, 87
self,
'idea' of, 55
identity of, 25–7, 27n, 28–9, 32
self-consciousness, first person and, 24–6, 31, 36
'self-feeling', 34
self knowledge, 34
sensation,
implications of, 60–1
intentionality of, 3–20, 61
language and, 62
memory and, 108, 120, 123
and perception, 64
of position, 71–4
subjectivity of, 44–56
sense-data, 3, 11, 13
perception and, 65, 67–8
sense-impression, viii, 11, 15, 107
and subjectivity of sensation, 46, 48–9, 54
sense-perception, 11
sensible properties, 39
'sensory core', perception and, 64
'sensory deprivation', 31, 34
sensory qualities, 48
shape, 39
sight,
colour and, 3
double vision, 67–8
first person and, 31
perception and, 66–8
significant connection, and grammar of "enjoy", 96–7
simultaneity, 183
size, 16, 39
Smullyan, Arthur, 176, 177, 178
Smythies, Y., 224n
Socrates, 120
soul, 88n
sound, 'standing qualities', 47–8
space, 150
beginning of existence, 153–5, 157
'spatio-temporal continuity', 216–17
specification, first person, 23
Spinoza, Baruch, 34n, 134
standard, and subjectivity of sensation, 51
'standing colours of objects', 46–7, 50, 54
subject-expression, grammar of "enjoy", 96, 98, 99
subjectivity, of sensation, 44–56
substance, 37–43
sufficient condition, causation and, 135
Summa Theologica (Aquinas), 88n, 152

Tajtelbaum, Al., 221–3
taste, 'standing qualities', 48
temporal clauses, "before" and "after", 180–95
Theaetetus (Plato), 50, 120
theodicy, vii
Theory of Human Action, A (Goldman), 163n
thinking, *see also* thought
first person, 31, 35
Thomson, Judith Jarvis, 214, 214n, 215
thought, *see also* thinking
expression of, 62–3
intention and, 59–61
Thucydides, 22n
time, 214, 219
beginning of existence, 148–62
quantification over, 185–6, 187, 194
reality of past, 103–19
Tractatus (Wittgenstein), 94
Treatise of Human Nature (Hume), 134, 139, 152
truth, 158

truth-functional connectives, 95, 96, 100, 174
truth-functionality,
extensionality and, 174, 178
of present-tensed proposition, 181
and subjunctive conditionals, 196-200
truth values, 168n, 176, 187
of subjunctive clause, 197, 198, 199, 200

"under a description", 208–19
understanding, of past, 106–7
universality, 135, 136
use, cf. mention, 220

validity, of reason, 226–8, 231
verb,
achievement verb, 192
grammar of "enjoy", 97
intentional, 11
psychological, 60
of sense, 12, 13
voluntary movement, 171

Waterman, David, 168n
will, freedom of, 145, 146
and soft determinism, 163, 165–8
Williams, Prof. Bernard, 9n
Wittgenstein, Ludwig, viii–ix, 70, 76, 114, 114n, 221
asymmetry, 60
imagination, 83
meaning, 118
pretending, 91
subjectivity of sensation, 52
Wittgenstein's Logical Positivism (Griffin), 38n
word, expressing thought, 59, 62
Word and Object (Quine), 16n, 173
Wright, G. H. von, 196